A genetic dissection of domestication-related traits in pea

by

Owen Williams

School of Biological Sciences

Submitted in fulfilment of the requirements for the Doctor of Philosophy

University of Tasmania December 2018

Author declarations

Declaration of originality

"This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief no material previously published or written by another person except where due acknowledgement is made in the text of the thesis, nor does the thesis contain any material that infringes copyright."

Authority of Access

This thesis may be made available for loan. Copying and communication of any part of this thesis is prohibited for two years from the date this statement was signed; after that time limited copying and communication is permitted in accordance with the Copyright Act 1968.

_____

Owen Williams

20th November 2018

Acknowledgements

I would like to take this opportunity to thank all those that have encouraged and supported me at the start and throughout this PhD. I would first like to express my sincere gratitude to my supervisors Jim Weller and Valérie Hecht for giving me this exciting opportunity with which I have thoroughly enjoyed these past 4 years. While these few sentences could never convey my appreciation, I am indebted for all their kind support, guidance and extensive knowledge. I would also like to give a special thanks to Jackie Vander Schoor who has played a significant a role in setting up this project and supporting me throughout as well as Raul Ortega who has been too kind for his own good and Kirstin Proft for solving many of my *R* based woes. I would also like to thank to Bea Contreras, Kelsey Picard and Andy Rubenach from the flowering time group and Jak Butler and Anicee Lombal from our office for their friendship and support (and patience).

I would like to thank all my friends that have made my experience in Tasmania so memorable and once or twice forgettable. At their own expense, they have made these past years some of the most exciting, funny and all round enjoyable, a time I shall look back on with a smile.

Thank you to all PhD candidates and staff at the School of Biological Sciences for providing a friendly and welcoming environment. A special thanks to Michelle Lang and Tracey Winterbottom for keeping my plants alive, Dr Jules Freeman and (now Dr) Jak Butler for their help with the construction of my linkage map. I would also like to thank Adam Smolenski and Sharee McCammon for their technical support and advice in the laboratory.

Closer to home, I would like to thank my family for their blind optimism, encouragement and support in all aspects of my time in Australia.

Table of contents

List of Figures

List of Tables

Abstract

Crop domestication refers to the process in which wild plants have become adapted for agricultural purposes. In different crops, common selective pressures unique to cultivation environments or agricultural practises have led to the accumulation of similar sets of traits, known as the domestication syndrome. Understanding the underlying genetic basis for domestication syndrome traits has been of interest for both plant breeders and evolutionary geneticists, as it can provide a window into the evolutionary history of a trait, lineage or crop species and assist in crop improvement efforts.

This study investigated the genetic control of domestication and diversification in pea (*Pisum sativum* L.); an important global legume crop and prominent genetic model system. As in other crop legumes, major phenotypic changes during this transition have included loss of seed dormancy and pod shattering, both considered to be critical domestication traits, as well as other changes such as earlier flowering, reduced branching and development of a more robust growth habit, which have been considered diversification or crop improvement traits. This study examined the genetic control of several of these traits in a biparental cross between the domesticated *P. sativum var sativum* line NGB5839 (a near-isogenic derivative of cultivar Torsdag) and the wild *P. sativum var humile* line JI1794*,* a representative of the "northern humile" lineage which is considered a major contributor to the modern day domesticated pea. It consisted of the generation of a high-density linkage map, QTL mapping, and the evaluation of specific QTL regions for dormancy-related and flowering time traits.

In **Chapter 3** a high density, high confidence genetic linkage map was developed for a population of 137 F_8+ Recombinant Inbred Lines (RILs) derived from the JI1794 x NGB5839 cross. This map incorporated 4599 DArT markers spanning a total of 1617cM. This exhibited good coverage, with an average of roughly three markers per cM with no markers >10cM apart. Map assembly was validated and assessed by a detailed examination of synteny with the well-characterized *Medicago truncatula* genome, and with four other legume genomes: *Lens culinaris*, *Trifolium pratensis*, *Cicer arietinum*, and *Phaseolus vulgaris*. This is the first study to compare the syntenic relationship of pea with recently sequenced *T. pratensis* and *L. culinaris. L. culinaris* is currently the most closely related species to pea for which a genome sequence is publicly accessible, and the high level of synteny makes this a useful new reference point for pea genetics.

The resulting map was used in **chapter 4** as the basis for QTL analysis of a range of domestication-related and other traits in the RIL population when grown under extended natural long-day conditions. Most traits were found to show relatively simple genetic control and, with only a few exceptions,

conditioned by one or two major loci (PEV ≥15%) and several minor loci (PEV ≤ 15%). QTL clustering was observed in four genomic regions near *LF*, *Hr*, *LE* and a region on LGVI, possibly in part reflecting pleiotropic effects of flowering time genes on other traits such as development rate and growth habit. This chapter provides new insight into the basis for pea domestication and subsequent diversification episodes and identifies major loci for dormancy and flowering time.

Chapters 5 and 6 explore in more detail the genetic control of several traits potentially related to seed dormancy. QTL analysis for permeability identified two previously unreported loci; a major locus on LGII (*Perm2*) and a minor locus on LGVII (*Perm7*). *Perm2* mapped close to major testa thickness QTL *TT2* and the well-known Mendel *A* locus.

In **Chapter 5**, the effect of *Perm2* and *TT2* were validated in segregating progenies derived from the JI1794 x NGB5839 F_2 population used for the generation of RILs, which were confirmed to have strong linkage to the *A* loci. The widespread existence of free-germinating pigmented domesticated lines suggests that the major change in permeability during domestication may not be caused by the *A* gene itself. Mutants for the *A*-interacting protein *A2* showed reduced testa thickness but no difference in permeability. This indicating that reduced testa thickness, may depend on biosynthetic pathways regulated by the *A*/*A2* complex however permeability remains inconclusive.

Loss of seed coat roughness and its co-ordinated emergence with the domesticated pea has previously been understood to implicate it in its control of seed dormancy. A major locus for roughness corresponding to the *GTY* locus was detected on LGVI, while a second novel QTL was identified on LGVII. **Chapter 6** investigated the genetic control of seed coat roughness in more detail and explored its relationship with testa thickness and permeability. Using advanced generation segregating populations, *GTY* was confirmed to co-segregate with increased testa thickness and was fine mapped to a region inferred to contain around 50 genes based on a pea transcriptome linkage map. A second testa thickness QTL in this region which had not been observed in the original QTL analysis was identified and shown to be closely linked to but genetically independent of *GTY*. Permeability experiments showed seed coat roughness and testa thickness on LGVI had no effect on hardseededness. This has been the first time the effect of testa roughness and thickness on permeability has been tested using advanced segregating populations and has provided the most detailed mapping to date of the *GTY* locus.

Chapter 7 explored in more detail the genetic control of newly identified flowering time QTLs on LGII and LGV (*DTF2* and *DTF5*). Like many other species, domesticated peas in general flower earlier with reduced photoperiod sensitivity. Among five long-day flowering time QTLs identified in Chapter 4, two

(*DTF3a* and *DTF6*) had previously been identified in short-day conditions in F_2 progeny of the same cross, while the third (*DTF3b*) mapped to the region of Mendel's *Le*. Using advanced generation segregating populations and expression analysis, floral inhibitory gene *LATE FLOWERING* (*LF*) was considered a likely candidate for *DTF2*, while *DTF5* mapped to a region known to include the *FTa1-FTa2-FTc* cluster of florigen genes. Plants carrying the NGB5839 alleles at *DTF2* and *DTF5* showed elevated expression of both *LF* and *FTa1* genes respectively and is consistent with the direction of the respective QTL effects, therefore suggesting that *DTF2* and *DTF5* might represent gain-of function variants in NGB5839. Involvement of the LGV *FT* gene cluster in natural variation for flowering time has not previously been documented.

Overall the results in this thesis make a significant contribution to our understanding to domestication and diversification of pea, and significantly extend previous studies on the genetic control of traits related to seed dormancy and flowering. In addition, the generation of a high-quality map will provide an ongoing resource for future analyses of a wider range of other domestication-related traits.

Chapter 1: General Introduction

1.1 Domestication

Crop domestication is a human induced process in which desirable variation for production and consumption has been consciously or unconsciously positively selected from the wild population. This has effectively resulted in an altered evolutionary pathway independent from the founder wild gene pool (Meyer and Purugganan, 2013) resulting in distinct morphological and physiological differences. Across diverse range of crops, a common suite of traits are often found that distinguish domesticated material from its wild ancestor. This has been referred to as the "Domestication Syndrome" (Harlan et al., 1973, Hammer, 1984, Burger et al., 2008). These features include loss of seed dispersal and dormancy, early flowering time and relaxation of environmental requirements for flowering induction, self-fertilisation, enlargement of edible organs with improved palatability and reduced toxicity, as well as modifications to plant architecture for improved harvest efficiency, such as determinate growth, apical dominance and reduced branching (Doebley et al., 2006, Meyer and Purugganan, 2013, Olsen and Wendel, 2013a).

Within this wider group of traits seed dispersal and dormancy are considered to have been essential at the earliest stages of cultivation, as they permit uniform germination and efficient harvesting (Olsen and Wendel, 2013b). In this sense they can be considered "true" domestication traits, and have been followed by other modifications that are not essential for effective intensive cultivation but have improved and extended the crop for human use - the so-called improvement or diversification traits. Unlike domestication traits, which can be best accessed through comparisons between wild and land-race material, improvement and diversification traits are best studied by comparing landraces and modern cultivars. These are distinguished from the domestication traits by the fact that they may vary between different domesticated lineages and are often related to productivity and adaptation in specific environments (Purugganan and Fuller, 2009, Olsen and Wendel, 2013a).

There has been much debate about the distinction between domestication and diversification traits, which is not particularly clear when applied across different crops with different evolutionary trajectories (Purugganan and Fuller, 2009, Olsen and Wendel, 2013b). In an attempt to clarify thinking on the topic, Abbo et al. (2014) proposed the concept of "crucial domestication traits" and suggested that these must have arisen relatively rapidly in distinct episodes. In this model, *domestication traits* were characterised as essential modifications for cultivation, adopted over a relatively short time period during the early or pristine domestication phase. This is proposed to result in qualitative differences that are ubiquitously expressed across the domesticated germplasm and might be expected to show simple genetic control. In legumes such as pea and lentil, such traits would include

loss of seed dormancy and seed dispersal mechanisms (Ladizinsky, 1985, Ladizinsky, 1987, Abbo et al., 2011). (Abbo et al., 2014) also distinguishes a later *diversification* phase which involves incremental (quantitative) adaptive modifications to non-essential traits such as seed size which increases only gradually over the course of crop evolution (Westoby et al., 1996, Purugganan and Fuller, 2009) and therefore more likely to show polygenic control. It should be noted that the importance of any given trait may also differ between crops, for example seed or fruit retention, previously described as a key domestication trait (Dong et al., 2014) is generally considered to be incorporated during the early domestication episode (Haberer and Mayer, 2015, Li and Olsen, 2016), however Fuller et al. (2014) reports asynchronous adoption with rapid evolution in wheat and barley but protracted evolution in rice.

This debate will be difficult to resolve and will depend on understanding the genetic differences that underlie the trait variation, and on comparisons of molecular evolution in the causal genes. The first step in this process is genetic definition and molecular identification of these causal genes.

1.2 Molecular assessment to the domestication episode

Genetic approaches are a powerful tool to unravelling key evolutionary episodes. As discussed above domestication syndrome traits may in most cases primarily show polygenic control (Stetter et al., 2017), and different genes contributing to regulation of the same trait can encounter varied selective intensities (Li et al., 2014b). This has created conflicts when defining domestication and diversification episodes, as genes can exhibit different evolutionary time lines (Meyer and Purugganan, 2013). For example the critical domestication traits, seed dormancy and indehiscent pods, should by their definition have distinct phenotype to their wild progenitor (Abbo et al., 2014), however, in legumes genetic evaluation has shown polygenic control (Weeden et al., 2002, Weeden, 2007), indicating quantitative variation. Therefore, domestication and diversification episodes can only be defined at a molecular rather than trait level. Using pod indehiscence in pea as an example, four QTLs have been recognised (Blixt, 1972, Weeden et al., 2002, Weeden, 2007) but only the major *Dpo1* locus has been consistently found in all wild x domesticated comparative studies (Weeden et al., 2002, Weeden, 2007). This indicates the *dpo1* mutation occurred prior to the divergence of the different landrace and cultivar varieties and therefore suggesting its nature as a critical domestication trait. In contrast the other loci identified were not universally present, pointing to their later adoption and a less critical function, consistent with their generally having smaller phenotypic effect than *Dpo1* (Weeden et al., 2002).

Population level genomic analysis involves isolating causative genes and comparing allele distributions across the germplasm, which can provide a window into its evolutionary history. For example, in maize

a mutant allele of the *TEOSINTE BRANCHING1 (Tb1)* gene confers increased apical dominance and reduced branching, which has been recognised a distinctive trait distinguishing it from its wild ancestor teosinte (Doebley et al., 1995, Doebley et al., 1997). The *tb1* mutation has been located within a 58—69kb upstream non-coding region which incorporates the *cis* regulatory element (Clark et al., 2006) and comparative genetic analysis between wild and domesticated forms demonstrated that the difference could be attributed to the insertion of two transposable elements (TE) insertions in this region (Zhou et al., 2011). Population level genomic analysis shows these TE insertions were present within all domesticated lines, as well as the closest wild subspecies *parviglumis* and *mexicana*, thereby bridging the evolutionary gap between the wild and domesticated forms. Other examples include origin of the loci *fw2.2* associated with increased fruit size in domesticated tomatoes, which occurred prior to domestication (Nesbitt and Tanksley, 2002) and in wheat, with the non-responsive flowering time *ppd-h1* gene originating from a wild accession in the Fertile Crescent, enabling its distribution to further latitudes (Jones et al., 2008).

Molecular analysis can be transferable across other crop models. Morphophysiological parallelisms between different crops often occur via convergent targeting of the same genes or pathways (Paterson et al., 1995). For example orthologs of *tb1* originally identified in maize has also been found within pearl millet (Remigereau et al., 2011), barley (Ramsay et al., 2011) sorghum (Mace et al., 2013) and rice (Takeda et al., 2003). Among different crops, similar phenotypic effects caused by similar genetic variations suggests highly conserved pathways and or limited genes for desired phenotypic outcome. However, examples such as in pearl millet where the *tb1* ortholog showed a more modest effect to branching than maize illustrates modifications within the regulatory pathways, therefore requiring additional mutations to achieve desired phenotype (Remigereau et al., 2011). There are also examples where similar phenotypes arise from independent genetic control. For example, in the Brassicaceae family, vernalisation involves the repression of floral repressor *FLORAL REPRESSOR CONSTANS* (*FLC*) via the *VERNALISATION2-polycomb* group repressive complex 2 (*VRN2-PRC2*). In other crop models such as legumes which lack the *FLC* gene, vernalisation is regulated via an alternative pathway (Alonso-Blanco et al., 2009, Jaudal et al., 2016).

1.3 Common molecular approaches

Domestication studies have traditionally been based on morphological analysis, however as previously discussed these studies are now old-fashioned and can be subjective and limiting in their nature. With the rapid technological advancements such as Next Generation Sequencing (NGS), molecular analysis has become cheaper and more accessible, subsequently becoming a normal analytical tool with domestication studies.

Domestication has the inevitable consequence of genetic diversity loss (Hyten et al., 2006, Gepts, 2014) due to limited lines carrying desirable traits being bred. This has had profound effects on the genomic architecture, resulting in molecular footprints which can be identified using analytical techniques. The Genome Wide Association Study (GWAS) approach analyses allelic diversity across a diverse germplasm collection and can quantify genetic diversity as selective sweeps. A common feature of these studies is the dramatic reduction of diversity during the initial domestication phase, as seen between wild and landraces populations, but a more modest reduction in the later stages between landrace and modern cultivar (Hufford et al., 2012, Cavanagh et al., 2013, Li et al., 2013). Despite a genome wide reduction in diversity loss, specific loci encounter more intense selective sweeps, likely indicating a desirable trait. Linkage disequilibrium studies which identify these selective signatures (Kim and Nielsen, 2004) are used to infer positions of domestication-related QTLs.

An alternative approach uses segregating populations to infer the genetic basis for traits. Quantitative Trait Locus (QTL) analysis has been a widely used technique in domestication studies, as this can model complex genotype – phenotype associations at both a trait (Zhu et al., 2017, Benech-Arnold and Rodríguez, 2018) and genome level (Koinange et al., 1996, Weeden, 2007, Lo et al., 2018). This method usually involves a biparental cross between two inbred lines with diverging phenotypes, which is then selfed to form genetically diverse, segregating population, which may be directly phenotyped or used to develop an inbred line population by single seed descent. Traits are measured across the population and QTLs identified by trait marker correlations.

With the advent of fully sequenced genomes, crop domestication and evolution research has accelerated. In 2002 rice, *Oryza sativa L. ssp. japonica*, was the first fully sequenced crop genome (Goff et al., 2002). Since then more than 25 different crop genomes have been sequenced with their genes annotated (Geleta and Ortiz, 2016). The availability of genome sequences has provided assistance in identifying causative genes and functional elements through candidate gene searches and the development of molecular markers, as well as providing the genomic tools and platforms for gene mapping, isolation and molecular breeding (Varshney et al., 2013a). In the absence of a fully sequenced genome, genetic linkage maps are developed and synteny with closely related species are made. In pea, a fully sequenced genome is not currently available, but the most recent mapping studies rely on a dense transcriptome consensus map and synteny with closely related species such as *M. truncatula* and *C. arietinum* (Tayeh et al., 2015). Although these species generally show strong syntenic relationships, these can be limiting particularly in poor syntenic regions such as on PsLGVI (Kaló et al., 2004, Bordat et al., 2011, Tayeh et al., 2015).

1.4 Evolution and importance of the domesticated pea

Over 2500 plant species have undergone domestication, with the earliest examples occurring within the Neolithic period over 10,000 years ago (Meyer et al., 2012, Zohary et al., 2012). Based on archaeobotanical and carbon dating, the cereal crops emmer and einkorn wheat (*Triticum turgidum subsp dicoccum* and *T. turgidum subsp monococcum*) and barley (*Hordeum vulgare)* were recognised as the earliest domesticated crops (Lev-Yadun et al., 2000). This occurred either in a core region in the South-Eastern Turkey/Northern Syria (Lev-Yadun et al., 2000) or across numerous independent sites within the Fertile Crescent (Fuller et al., 2011). Around this time four legume crops were also domesticated in this region including lentil (*Lens culinaris*) 10,100-9700 years ago and pea (*Pisum sativum*), bitter vetch (*Vicia ervilia*) and chickpea (*Cicer arietinum*) 9,900 – 9,500 years ago. Despite the fact that this archaeobotanical evidence is not quite as old as the cereals, it is generally thought that legume domestication occurred either alongside or closely following cereals (Lev-Yadun et al., 2000, Zohary et al., 2012).

With increase in global population and rising standard of living, it is estimated that a 70% increase in food production will be required by 2050. It is expected that legumes will play an significant role in bridging this gap due to their high protein and nutrient content (Duranti and Gius, 1997, Varshney et al., 2013b). Economically the legumes represent the second most important staple food group behind cereals, with pea having the second highest worldwide production in pulse crops (14.4Mt) (FAOSTAT, 2018). Pea also has other important roles in farming practises, including providing fodder for livestock and maintaining fertile soils by sequestering atmospheric nitrogen. An investigation of the underlying genetic components of its domestication is therefore of value, from both an economic and a scientific perspective, as it would provide further understanding of its evolution and offer a platform for future crop improvement efforts.

1.4.1 Taxonomy and Domestication of Pea

The *Pisum* germplasm has previously been classified into three main groups *P. fulvum*, *P. sativum var. abyssinicum* and *P. sativum spp* (Ellis et al., 1998) with *P. sativum var. sativum, P. sativum var. humile*, *P. sativum var. elatius*, *P. sativum var. jomardii*, and other minor taxa grouped together as *P. sativum* subspecies (Vershinin et al., 2003, Jing et al., 2010, Smýkal et al., 2011). While *P. fulvum* is clearly a valid species, as both molecular analysis (Vershinin et al., 2003) and reproductive compatibility studies show it is unequivocally distinct from all other lineages (Maxted and Bennett, 2001). It has long been recognised that the cultivated pea comprises two independent lineages, including *P. sativum var. sativum* (comprising field, snap and snow peas) and the independently domesticated "abyssinicum" forms (Jing et al., 2010), but the relationship between them is not entirely clear.

Whilst phylogenetic P. sativum var abysiniccum has, by some been reffered to as a subspecies of *P. sativum* (Vershinin et al., 2003), a distinct chromosomal rearrangement between the two (Smýkal et al., 2011, Bogdanova et al., 2014) and a significant proportion of *P. fulvum* derived molecular markers arguably could signify a separate species. However if taxonomic rank was based on genetic divergence from founder population, then arguably this would place *P. sativum var. abyssinicum* as a subspecies, as certain geographically distinct *P. sativum* accessions such as the Central Asian "Afghanistan" types show similar levels of genetic isolation (Jing et al., 2012, Kwon et al., 2012, Holdsworth et al., 2017). This remains an ongoing debate which is not a focus of this study.

Genetic analysis indicates that *P. sativum var. sativum* is most likely to have originated from wild *P. sativum var. humile* (Zohary and Hopf, 1973, Hoey et al., 1996, Vershinin et al., 2003, Jing et al., 2010) while *P. sativum var*. *abyssinicum* shows a clear genetic contribution from *P. fulvum* (Jing et al., 2010, Weeden, 2018). As shown in Figure 1.1 (highlighted in light green), the geographic distribution of *P. sativum var. humile* is restricted to the Northeast Israel, Syria, south Turkey and the western side of the Zagros mountains in Iran (Zohary and Hopf, 1973). This likely represents the origin of the domesticated *P. sativum var. sativum* and includes the earliest archaeological sites in Turkey (Çayönü) and Syria (Bouqras) where smooth seededness, an indicator of domestication in pea, has been documented (Zohary and Hopf, 1973, Zohary et al., 2012). The domesticated pea is thought to have spread throughout Southern Eurasia, where it is assumed to have diverged resulting in two distinct lineages (Jing et al., 2010); an eastern expansion towards the Indian subcontinent and Himalayan region (highlighted in dark green) giving rise to the *Afghanistan* germplasm group, and the more prominent western expansion to Mediterranean Europe (highlighted in orange) which eventually gave rise to modern *P. sativum var. sativum* cultivar. Although the eastern and western spread occurred relatively rapidly, a strong long-day photoperiod requirement for flowering prevented its movement into more northern latitudes (Purugganan and Fuller, 2009) (highlighted in light blue). This was eventually overcome by mutation in the flowering time gene *HIGH RESPONSE* (*HR*)*/ EARLY FLOWERING 3a (ELF3a),* resulting in early flowering in short photoperiods and allowed a transition from winter to spring cropping (Weller et al., 2012). The fact that both *P. sativum var. abyssinicum* and *the Afghanistan* domesticated *var. sativum* germplasm carry a functional *ELF3* gene further supports the idea that this transition occurred from a western Mediterranean genepool.

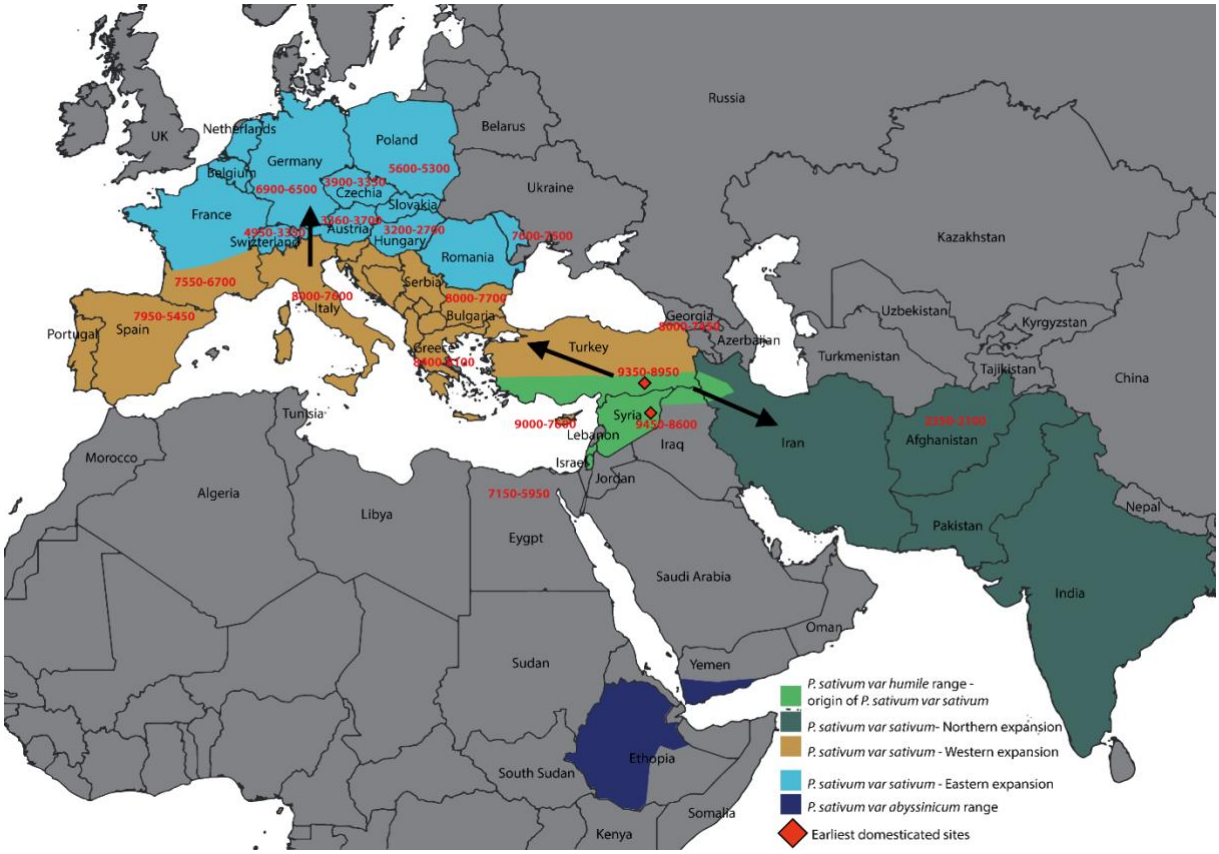


Figure 1. 1: Distribution and expansion of wild (P. sativum var. humile) and domesticated pea (P. sativum var. sativum and P. sativum var. abyssinicum) based from archaeobotanical Zohary et al. (2012) and genetic evaluation (Jing et al., 2010, Weller et al., 2012). Arrows indicate expansion direction, values in red denote earliest archaeological findings according to Zohary et al. (2012). Not all P. sativum var. sativum regions highlighted.

As shown in Figure 1.1, the distinct domesticate *P. sativum var. abyssinicum* is restricted to an isolated region in the Ethiopian highland and Southern Yemen region (highlighted in dark blue). Geographical barriers have prevented introgression with *P. sativum var. sativum* which has resulted in a genetically distinct but much narrower genepool. As previously mentioned, genetic analysis has indicated a greater contribution from the *P. fulvum*, with some authors suggesting its origin may have arisen from a *P. sativum* x *P. fulvum* cross (Vershinin et al., 2003, Kosterin, 2017). It was proposed this originated from a similar region as *P. sativum var. sativum*, in the Western Fertile Crescent region and was then transported and developed in North East Africa (Jing et al., 2010). However based on limited shared alleles between *P. fulvum,* Weeden (2018) argues *P. sativum var. abyssinicum* would not have occurred from a *P. sativum* x *P. fulvum* hybridization in the last 10,000 years. Instead Weeden (2018) suggest *P. sativum var. abyssinicum* more likely originated from a *P. sativum var. elatius* x *P. sativum var. sativum* hybridisation (Kwon et al., 2012).

In a comparison of molecular control of key domestication traits, Weeden (2018) proposed that these two domesticated varieties diverged before the evolution of pod indehiscence but after the loss of seed dormancy. This was based on flanking genes to the pod indehiscent *Dpo1* locus showing closer

similarity to *P. sativum var. elatius* than domesticated *P. sativum var. sativum*, implying independent evolution, whereas for the proposed seed dormancy QTL on LGIII, flanking genes *Rms1 and Rb* showed closer similarity to *P. sativum var. sativum,* therefore suggesting a common origin. It should be noted this study did not directly assess seed dormancy, but instead measured the reduction on testa thickness, on the assumption that this is closely related to the loss of dormancy that occurred during domestication, therefore it is arguable whether this embodies the seed dormancy region.

In this study there will be a focus on understanding the genetic control of seed dormancy related traits and flowering time. These traits represent two important evolutionary phases, including the origin of domesticated lineage by loss of seed dormancy mechanisms and its expansion in eco-geographical range by modifications in flowering time habit.

1.5 Seed Dormancy

Seed dormancy is the phenomenon in which seeds are not able to germinate immediately upon reaching maturity. Typically found within wild populations, seed dormancy allows the and persistence of seeds within the soil bank and results in a wide temporal spread of germination (Matilla et al., 2005). It is effectively a bet-hedging strategy against catastrophic loss due to adverse environmental conditions (Cohen, 1966, Bulmer, 1984), and ensures that germination can be restricted to a time that environmental conditions are favourable (Finkelstein et al., 2008). Conversely, in agronomic environments where crops are grown in relatively uniform conditions, synchronous germination is an advantage and dormancy is counterproductive to yield (Lenser and Theißen, 2013). As previously mentioned, loss of seed dormancy has been reported as fundamental for the domestication of legume species, with experimental field trials reporting net yield loss when cultivating dormant seeds (Abbo et al., 2011).

1.5.1 Mechanisms of legume seed dormancy - physical

The transition from dormant state to germination begins with the influx of water and is completed by the elongation of the embryonic axis. The maintenance and release of seed dormancy for germination in optimal conditions requires the fine balancing of environmental and physiological inputs. These complex inputs can be categorised into two mechanistic pathways, physical and endogenous dormancy (Bewley, 1997).

Physical seed dormancy also referred to the dormancy that results from the inability of water to permeate the seed coat is also referred to as "hardseededness". It is present in at least 17 different families (Baskin and Baskin, 2000), prominently including the legumes. The mechanisms that regulate legume seed dormancy are poorly understood (Smykal et al., 2014), even more so when considered

from a domestication perspective. There are major phenotypic disparities that distinguish the wild and domesticated seeds, which could be affecting permeability properties (Figure 1.2). These can be subdivided into structural which includes testa thickness and seed coat roughness (Harris, 1987, Baskin et al., 2000) and chemical properties such as pigmentation and fatty acid composition (Shao et al., 2007, Zhou et al., 2010, Cechová et al., 2017). Not surprisingly, physical seed dormancy has been recognised as a quantitative trait (Foley, 2001, Finch-Savage and Footitt, 2012, Smykal et al., 2014, Nakamura et al., 2017), although it is not clear how this can be reconciled with its loss during domestication.

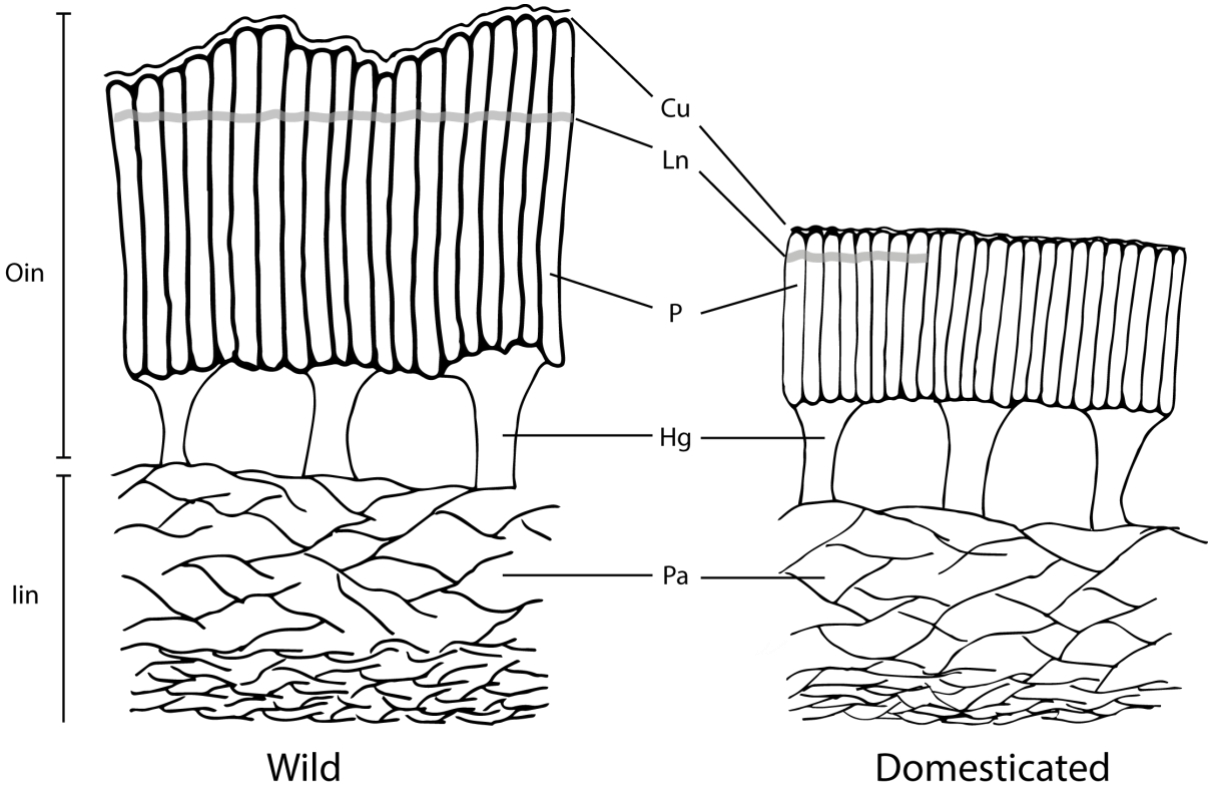


Figure 1. 2: Schematic representation of the mature wild and domesticated pea seed coat structure based from transverse sections in (Lush and Evans, 1980, Smykal et al., 2014). The pea seed coat consists of three layers i) palisade (P) including Cutin (Cu) and light line (Ln), ii) Hourglass (Hg), and iii) Parenchyma (Pa). The P and Hg layers form the outer integument (Oin) and Pa forming part of the inner integuments (Iin) which also includes the endosperm (not included in diagram).

1.5.1.1 Structural components

In some species specialised regions within the seed coat are known to provide natural openings that regulate permeability, such as the hilum, raphe and micropyle (Lush and Evans, 1980), however the universal significance of these structures has been debated (Ma et al., 2004b). Other components within the palisade layer have been associated with dormancy, including presence of a light-line (resulting from variation in refractive index in macrosclereid layer), thickening of the cuticle and inner tangential and radial cell walls, seed coat roughness, surface deposits and reduction in frequency and size of the pits (Figure 1.2). However, the most striking structural variation is the overall testa thickness

(Weeden, 2007, Smykal et al., 2014, Hradilová et al., 2017) with reductions of over 50% (Plitmann and Kislev, 1989) in domesticated relative to wild forms. This is consistent with archaeological evidence of recurrent thinning in seed coat thickness in relation to domestication stage (Smith, 1997, Smith, 2006, Murphy and Fuller, 2017), implying a continued selective pressure throughout the domestication process.

A strong link between testa thickness and loss of dormancy has often been assumed due to the concurrent thinning and increased and permeability in domesticated lines (Harlan, 1992, Argel and Paton, 1999, Baskin, 2003, Maass and Usongo, 2007). With few exceptions, this trend appears universal across the legume species (Lush and Evans, 1980, Lush and Evans, 1981, Maass, 2006, Murphy and Fuller, 2017), particularly with respect to thinning of the palisade layer (Werker et al., 1979, Lush and Evans, 1981, Plitmann and Kislev, 1989, Miao et al., 2001). Temperature, rainfall and humidity are important environmental factors that facilitate the breaking of physical seed dormancy (Long Rowena L. et al., 2015), which is generally promoted by temperature fluctuations over extended periods. (McKeon and Mott, 1982, Baskin, 2003). This is thought to cause expansion and contraction of the seed coat that eventually exposes faults allowing water to penetrate across the impermeable barrier. Therefore a thicker seed coat would likely offer a greater mechanical resistance to imbibition (Ojomo, 1972, Debeaujon et al., 2000). Nevertheless, reports of permeable but thick testa and impermeable, thin testa seeds (Lush and Evans, 1980, Hradilová et al., 2017) indicates that testa thickness cannot fully explain seed dormancy (Souza and Marcos-Filho, 2001).

1.5.1.2 Chemical components

Plant cells have shown to regulate permeability via lignification, suberisation and cutinisation (Nawrath et al., 2013). Cutin and suberin are both glycerol-based aliphatic polyesters with embedded waxes, and can be found on the aerial surfaces and the outer integument layer of the seed coat respectively (Molina et al., 2008), whereas lignin is a complex aromatic heteropolymer found within the cell walls. While the role of lignin in seed dormancy is poorly understood, suberin and cutin have both been demonstrated key roles in permeability (Nawrath, 2006, Beisson et al., 2007). Suberin and cutin are comprised of a complex assortment of multiple lipids (Beisson et al., 2012) and share similar chemical and structural affinities, therefore indicating common regulatory mechanism to permeability. This is evident by mutant suberin and/or cutin biosynthesis lines showing increased permeability due to modifications to its chemical (Beisson et al., 2007, Chai et al., 2016, Gou et al., 2017) and structural (Nawrath, 2006, De Giorgi et al., 2015, Lashbrooke et al., 2016) composition.

1.5.1.2.1 Pigmentation

The diverse range of colours exhibited within flowers, fruits, seeds and leaves are the result of the wide assortment of pigments compounds, including chlorophyll, betalains, carotenoids and flavonoids (Tanaka et al., 2008), with the later often associated with seed dormancy. Flavonoids are hydroxylated phenolic compounds ultimately synthesised through the phenylalanine pathway which is the precursor to other biosynthesis pathways including lignin (Boerjan et al., 2003) and other important metabolites (Vogt, 2010). The flavonoid biosynthesis pathway is ubiquitous across the plant kingdom, with over 10,000 structures identified (Brunetti et al., 2013), and this diversity alludes to a wide range of biological functions within the plant (Ferrer et al., 2008). These include roles in stress responses (Petrussa et al., 2013), UV-B protection, pest and pathogen resistance (Agati and Tattini, 2010), pollinator attraction (Kevan et al., 1996), pollen fertility (Shirley, 1996) hormone signalling and nodulation (Hirsch, 1992) as well as increasing seed dormancy and longevity (Shirley, 1998, Debeaujon et al., 2000). Flavonoids can be subdivided into six major groups based on their chemical structure; flavones, flavonols, flavanones, flavanols, anthocyanidins, and isoflavones, however pigmentation is only expressed in flavonols (pale yellow) anthocyanidins (red, blue, orange and yellow) and proanthocyanidins (brown) compounds (Díaz et al., 2010). In arabidopsis 54 pigment compounds have been isolated, including 35 flavonols, 11 anthocyanins and eight proanthocyanidins (Saito et al., 2013).

The synthesis of these compounds involves multiple enzymatic steps and involves numerous intermediate metabolites. As illustrated in Figure 1.3, Appelhagen et al. (2014) has defined the overall flavonoid pathway into 9 distinct phases which, based on the arabidopsis model comprises over 20 genes, as detailed in table 1.1. These can be separated into functional categories of biosynthesis (*TT3*, *TT4*, *TT5*, *TT6*, *TT7*, *TT10*, *BAN* and *LDOX*,) regulation (*TT1*, *TT2*, *TT8*, *TT16*, *GL2*, *TTG1* and *TTG2*) and transport (*TT9*, *TT12*, *TT13*, *AHA10* and *GST26*) genes.

The dedicated flavonoid biosynthesis pathway initiates with P-Coumaroyl-CoA, which is converted into dihydroflavonols (Flavonoid stages 1 to 4 – Figure 3) via the so-called Early Biosynthetic Genes (EBG) *CHS*, *CHI*, *F3H* and *F3'H*. At this stage, metabolism of dihydroflavonols can follow the anthocyanin/proanthocyanidin pathway (Flavonoid stages 5 to 9 – Figure 3) which involve the Late Biosynthesis Genes (LBG) or branch off to form flavonol compounds. Anthocyanins and proanthocyanidins (PA) form two competing branching pathways. Cyanidins can either be converted to anthocyanins by glycosylation, acylation and/or methylation (Shi and Xie, 2014) or converted to the PA precursor, epicatechin (Flavonoid stage 7) via anthocyanin reductase (*ANR*) (Xie et al., 2003). The regulation of the biosynthetic genes are controlled by *MYB-WD40-bHLH* (MBW) transcription factor complexes (Xu et al., 2015) of which the *TT2-TT8-TTG1* plays a central role (Nesi et al., 2000, Baudry

et al., 2004), although other MBW complexes; *MYB5–TT8–TTG1*, *TT2–EGL3*/*GL3–TTG1* can have overlapping functional roles (Xu et al., 2014, Xu et al., 2015).

Loss of pigmentation has commonly occurred with domestication or diversification (Plitmann and Kislev, 1989), potentially indicating an important agronomic function, with the possibility that one of its roles is regulating seed dormancy (Marbach and Mayer, 1974, Werker et al., 1979). The accumulation of proanthocyanidins within the palisade and parenchyma layers of the seed coat (Ferraro et al., 2014) has been associated with physical seed dormancy (Debeaujon et al., 2000), including in pea (Marbach and Mayer, 1974). This was highlighted by Zhou et al. (2010) who showed in soybean, that the content of epicatechin, a precursor to proanthocyanidin, correlated with seed coat permeability, whereas components within the competitive anthocyanidin pathway were down regulated. This was further supported by a permeability experiment conducted on 19 arabidopsis biosynthesis mutant lines (Debeaujon et al., 2000), where in general it can be seen mutations occurring further upstream had more severe effects on permeability (Table 1.1). This was particularly evident when comparing between EBG and LBG mutants, consistent with a proposed functional role of PA in seed dormancy.

*Table 1. 1: Table detailing genes involved in flavonoid biosynthesis pathway. *Permeability scored from 1 (most permeable) to 19 (least permeable) using data from Debeaujon et al. (2000) on arabidopsis mutant lines. **Flavonoid biosynthesis stage based on (Appelhagen et al., 2014)*

Mutant	Permeability score*	Flavonoid stage**	Gene/enzyme	Function	Role	Reference
ttg1	1	5 to 9	*WD40* (*A2*)	Part of the MBW ternary transcription factor regulating LBG	regulation	(Baudry et al., 2004)
tt11 (tds4/tt17/ tt18)	2	6	Leucoanthocyanidin dioxygenase (*LDOX*)	Leucocyanidin to cyanidin	Biosynthesis	
tt5	3	2	Chalcone isomerase (*CHI*)	Tetrahdroxchalcone to Naringenin	biosynthesis	
tt6	4	3	Flavanone 3-hydroxylase (*F3H*)	Naringenin to dihydrokaempferol	biosynthesis	
tt7	5	4	Flavanone 3' hydroxylase (*F3'H*)	Dihydrokaempferol to dihydroquercetin	Biosynthesis	(Hadas, 1976, Han et al., 2010)
tt2	6	5 to 9	*MYB123*	Part of the MBW ternary transcription factor regulating LBG	regulation	(Baudry et al., 2004)
tt4	7	1	Chalcone synthase (*CHS*)	4-coumaroyl-CoA and Malony-COA to tetrahdroxchalcone	biosynthesis	
tt3	8	5	Dihydroflavanol reductase (*DFR*)	Dihydroquercetin to leucocyanidin	biosynthesis	
ban	9	7	Anthocyanidin reductase (*ANR*)	Cyanidin to epicatechin	biosynthesis	(Xie et al., 2003)
tt14/tt19	10	8	Glutathione S-transferase 26 (*GST26*)	Epicatechin condensed to proanthocyanidin	Transport	(Gao et al., 2010) (Xu et al., 2014
tt13	11	8	Tonoplast P3A -TPase	Transport of flavonoids to vacuoles	Transport	(Appelhagen et al., 2015)
ats	12	Not involved	N/A	N/A		
tt1	13	5 to 9	WIP1 zinc finger	Interacts with *TT2* in the *TT2-TT8-TTG1* MBW complex to increase BAN activity	Regulation	(Azani et al., 2017)
ap2	14	Not involved	APETALA 2 (*AP2*)	N/A	N/A	
tt12	15	8	MATE family protein	Epicatechin condensed to proanthocyanidin	Transport	(Zhao and Dixon, 2009)
tt8	16	5 to 9	bHLH (Mendel *A*)	Part of the MBW ternary transcription factor regulating LBG	regulation	(Nesi et al., 2000)
tt9/GFS9	17	8		Transport of flavonoids to vacuoles	Transport	(Ichino et al., 2014)
gl2	18	5 to 9	N/A	Potential repressor of *MBW* components	Regulation	(Wang et al., 2015)
tt10	19	9	Laccase-like 15 (*LAC15*)	Proanthocyanidins to oxidised tannins	Biosynthesis	(Pourcel et al., 2005)
tt15	-	?	*UGT80B1*	unknown	unknown	(Xu et al., 2017)
aha10	-	8	Autoinhibited H+ ATPase 10 (*AHA10*)	Epicatechin condensed to proanthocyanidin	transport	
tt16	-	5 to 9	Agamous-like 32 (*AGL32*)	Regulation of *TT8*	Regulation	(Xu et al., 2013)
ttg2	-	9	Transparent testa glabra 1 (*TTG1*)	Regulation of *TT12* and *AHA10* expression	Regulation	(Gonzalez et al., 2016a)

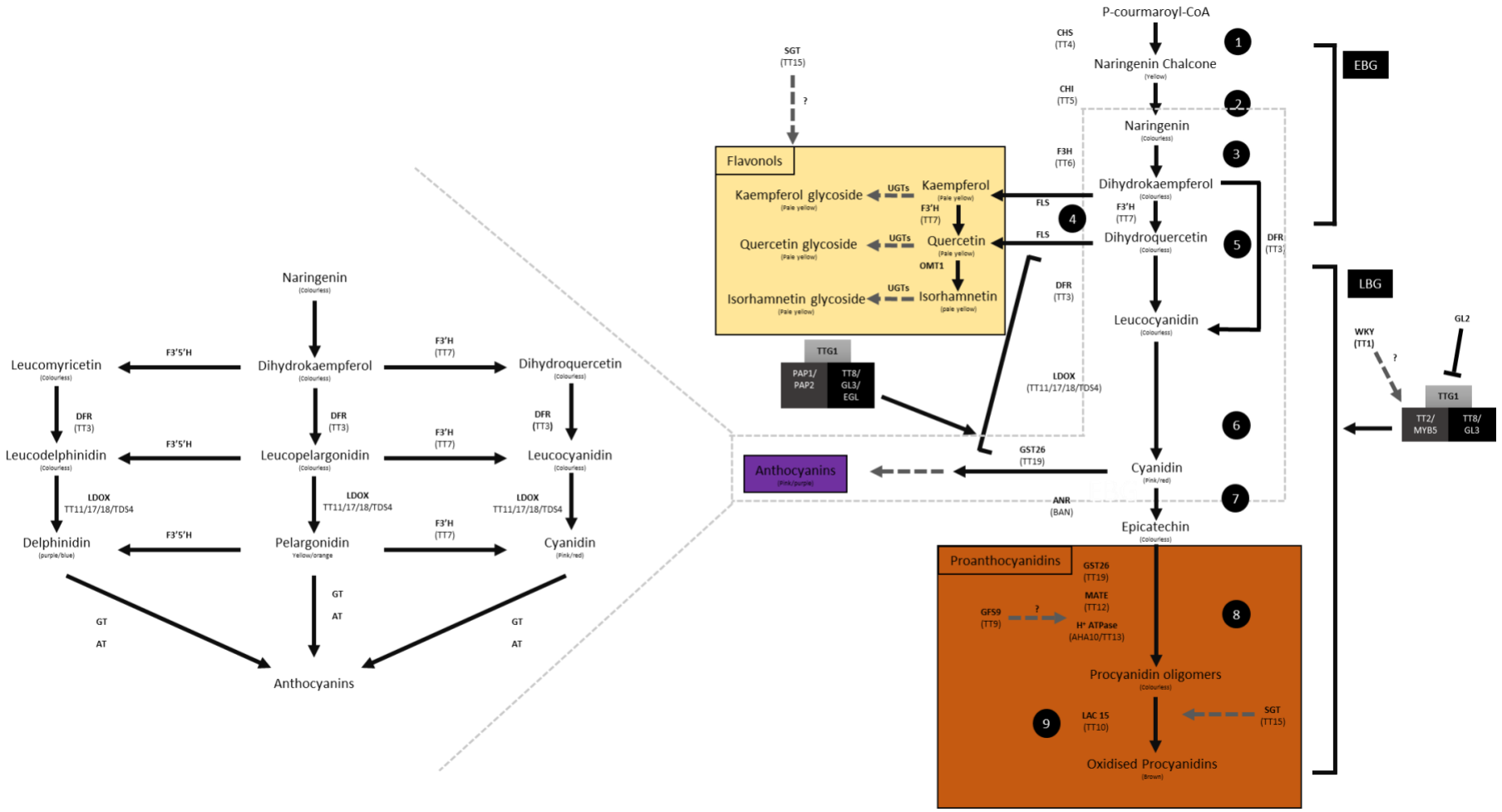


Figure 1. 3: Schematic overview of the flavonoid biosynthesis pathway in arabidopsis, illustrating the synthesis of Flavonols, Anthocyanins and Proanthocyanidins adapted from (Appelhagen et al., 2014, Smykal et al., 2014). Structural proteins and corresponding mutants are based from tables 1.1. The MBW regulatory complexes are based from (Xu et al., 2014) Colours of each compound based from (Lian et al., 2017). Acronyms for each enzyme are as follows: CHS, Chalcone Synthase; CHI, Chalcone Isomerase; F3H, Flavanone 3-Hydroxylase F3'H, Flavonoid 3′-Hydroxylase; FLS, Flavonol Synthase; OMT, O-Methyl Transferase; UGTS, Uridine Diphosphate Glycosyltransferase; DFR, Dihydroflavonol 4-Reductase; LDOX, Leucoanthocyanidin Dioxygenase; ANR, Anthocyanidin Reductase. ANS, Anthocyanidin Synthase; AT, Acetyltransferases; GT, Glycosyltransferases

1.5.2 Mechanisms of seed dormancy - Physiological

After the seed becomes imbibed, it is released from it quiescent state (Holdsworth et al., 2008, Weitbrecht et al., 2011) where it becomes metabolically active, but can still be maintained in a state of dormancy prior to sustained growth (Bewley, 1997, Han and Yang, 2015). This is known as endogenous seed dormancy and its control is accomplished by the integration of a diverse range of endogenous and environmental pathways. It is considered to be primarily modulated by a balance between two antagonistic hormones; abscisic acid (ABA) and gibberellin (GA) (Holdsworth et al., 2008) although other factors such as brassinosteroids, ethylene, reactive oxygen species and nitrogen containing compounds (Finkelstein et al., 2008) may participate to a lesser extent.

The onset and depth of dormancy is correlated with the accumulation of ABA, while ABA-deficient mutants exhibit non-dormant phenotypes. ABA is induced during the early to middle phase of embryo maturation (Chono et al., 2006), which coincides with its most dormant state. Germination sensitivity increases by the degradation of ABA over time (Millar et al., 2006) resulting in an age-dependent dormancy, a phenomenon known as “after-ripening”. The rate of ABA degradation varies between dormant and non-dormant phenotypes (Ali-Rachedi et al., 2004) and is influenced by environmental factors such as light, temperature (Lim et al., 2013) and other molecular components involved in the ubiquitin-mediated degradation pathway (Lopez-Molina et al., 2001, Zhang et al., 2005) such as *DWA1* and *DWA2* (Lee et al., 2010).

GA is fundamental for germination, although its role in promoting germination is dependent on the concurrent down regulation of ABA (Gubler et al., 2005, Zhu et al., 2009), which is demonstrated by exogenous GA treatments unable to alleviate dormant seeds (Ali-Rachedi et al., 2004, Millar et al., 2006). GA stimulates germination by mobilising seed storage compounds and promoting embryo expansion by weakening of the endosperm or seed coat barriers (Groot and Karssen, 1987) which is normally maintained by ABA (Groot and Karssen, 1992). This demonstrates maintenance and breakage of dormancy is determined by ABA and GA metabolism, which regulate each other’s expression via a series of feedback loops (Finkelstein et al., 2008, Shu et al., 2016).

While endogenous seed dormancy has a critical role in many plant species such as arabidopsis, there have been very few reports in legumes (Van Staden et al., 1987) with no known reports in pea, which instead is strongly modulated by physical seed dormancy (Baskin and Baskin, 2007).

1.6 Flowering time

Flowering time is an important adaptive trait for ensuring reproductive success. In native populations this is primarily regulated by photoperiod and vernalisation mechanisms which delay flowering until

conditions are most favourable for seed set and maturation. From an agricultural perspective it is an important trait because the relaxation of these mechanisms enabled shorter growing seasons and production across a wider ecogeographical range (Faure et al., 2012, Weller et al., 2012, Peng et al., 2015). In obligate long-day plants, the relaxation in photoperiod sensitivity facilitated the shift from winter to spring cropping in temperate climes, by the shortening growth cycle (Orf et al., 1999, Gaur et al., 2015). This was evident by the expansion of agricultural crops such as cereals (Jones et al., 2008) and legumes (Weller et al., 2012) from the Fertile Crescent to diverse environments.

1.6.1 Molecular control of flowering

The molecular control of flowering time and signalling pathways have been well characterised, particularly in model species such as arabidopsis. In arabidopsis, flowering is controlled by a complex network of genetic and endogenous signalling pathways which converge to induce expression of floral integrator genes *FLOWERING LOCUS T (FT)* and *SUPRESSOR OF EXPRESSION OF CONSTANS1 (SOC1). FT* and *SOC1* in turn promote the expression of the floral meristem genes *APETALA 1 (AP1), LEAFY (LFY)* and *CAULIFLOWER (CAL)*, a pathway highly conserved among all flowering plants. In arabidopsis considerably more than 100 genes have been shown to affect flowering time, and six major genetic pathways for regulation of flowering time have been characterized; the photoperiod, vernalisation, gibberellin, temperature, autonomous and age pathways. In comparison, the discovery of flowering genes in legumes has not been as extensive with 4 loci identified in chickpea (Ortega, 2018), around 10 in soybean and around 20 in pea (Weller and Ortega, 2015).

1.6.1.1 Flowering Locus T in legumes -FT

FT is part of a gene family with similarities to animal phosphatidylethanolamine-binding proteins (PEBPs). In arabidopsis the wider *FT* family also includes *TERMINAL FLOWER1 (TFL1), TWIN SYSTER OF FT (TSF), ARABIDOPSIS THALIANA CENTRORADIALIS (ACT), BROTHER OF FT AND TFL1 (BFT)* and *MOTHER OF FT AND TFL1 (MFT)* (Danilevskaya et al., 2008). The FT protein has been recognised for its function as a universal floral signalling molecule “florigen” (Song et al., 2015). Whereas in arabidopsis, the *FT* subfamily is represented by only two, very similar genes, *FT* and *TWIN SISTER OF FT* (*TSF*), the *FT* gene family in legumes has undergone an expansion that appear to have a more diverse range of functions and expression patterns. Three functionally distinct *FT* clades have been characterised, namely *FTa*, *FTb* and *FTc*, which show distinct regulation patterns (Hecht et al., 2011, Laurie et al., 2011, Książkiewicz et al., 2016).

The number of PEBP genes in Medicago and pea was previously evaluated in Hecht et al. (2011) and (Laurie et al., 2011). These studies identified five *FT* genes in pea and Medicago, including two *FTa*

genes (FTa1 and FTa2), two *FTb* genes (FTb1 and FTb2) and one *FTc* gene. A recent evaluation across nine legume species by Książkiewicz et al. (2016) identified a novel *FTa* homolog within *M. truncatula* (Medtr6g033040) which was later identified within all Galegoid clade species (*M. truncatula* and *Cicer arietinum*) as well as *Glycine max* belonging to the *phaseloid* clade. Although this study did not include additional species from the galegoid clade, it is considered likely to be present within the *Pisum* genus which also belongs in the galegoid clade.

All identified *PsFT* genes have been shown to promote flowering to varying levels when overexpressed in an arabidopsis *ft-1* mutant, but there is evidence of subfunctionalisation. Grafting experiments by Hecht et al. (2011) indicated a key role for both *FTa1* and *FTb2* as independent mobile floral induction signals. It was proposed that *FTa1* is induced in the leaves in long-day and short-day conditions and is responsive to vernalisation treatment. Analysis of *FTa1* mutants suggests that the signal is transmitted from leaves to shoot apex where it suppresses floral repressor *LATE FLOWERING* (*LF*) expression while upregulating *FTc* and its own expression in a positive feedback loop (Hecht et al. 2011). Like *FTa1*, *FTb2* is also induced in the leaf, but in contrast, is only expressed in long-day conditions. *FTb2* is thought to regulate *FTa1* in the leaf and apex as well as inducing *FTc*. Conversely the expression of *FTc* is exclusively found within the shoot apical meristem and is thought to be transcriptional target of the *FTa* and *FTb* floral signals (Hecht et al., 2011, Weller and Ortega, 2015).

1.6.1.2 TFL1-like floral repressors in legumes

In model plant system arabidopsis, *TERMINAL FLOWER1 (TFL1)* is a repressor of flowering, working to maintain indeterminate inflorescence growth (Ratcliffe et al., 1998). This acts antagonistically to *FT* by suppressing floral identity genes (Boss et al., 2004, Hanzawa et al., 2005, Wickland and Hanzawa, 2015) to regulate floral development to optimal conditions (Kobayashi et al., 1999, Lifschitz et al., 2014, Kaneko-Suzuki et al., 2018).

Similar to *FT*, *TFL1* family has also undergone expansion and sub functionalisation within the legumes (Hecht et al., 2011, Laurie et al., 2011). Książkiewicz et al. (2016) subdivided these into three distinct subclades *TFL1a TFL1b* and *TFL1c*, which exhibit varying level of expansion within each subclade. In pea a single homolog was identified within each *TFL1* subclade (Hecht et al., 2011, Laurie et al., 2011), with *PsTFL1a* and *PsTFL1c* being encoded for *DETERMINATE (DET)* and *LATE FLOWERING (LF).* The *det* mutant does not appear to affect flowering time but is involved in determinate growth*,* whereas the *lf* mutants flower earlier than wild type but does not influence plant architecture (Murfet, 1975, Foucher et al., 2003, Sussmilch et al., 2015), whereas the role of *TFL1b* is currently unknown

1.6.1.3 Mechanisms controlling flowering time by seasonal cues in legumes

The timing of flowering is dependent on seasonal changes, including day length sensitivity and vernalisation. In many plant species flowering can be stimulated by the exposure of cold temperatures, a process known as vernalisation. In model system arabidopsis, this is mediated by *FLOWERING LOCUS C (FLC)* and *FRIGIDA (FRI)* (Caicedo et al., 2004), where *FLC* is highly expressed in non-vernalised plants and suppresses flowering by repressing floral promoting *FT, SOC1* and *FD* expression. After the prolonged cold period, *FLC* is stably downregulated in vernalised plants which promotes the floral induction cascade (Michaels, 2009).

Floral timing can also be highly sensitive to photoperiod, which depending on the plants requirements can be classified as either; long-day (LD), short-day (SD) and day-neutral. Many temperate species, such as pea, arabidopsis and barley show a LD photoperiod requirement. In arabidopsis, circadian clock genes drive rhythmic accumulation of CONSTANS (CO) protein, which is a direct transcriptional activator of *FT* (Simon and Coupland, 1996). In the first comprehensive analysis for flowering time related genes in legumes , Hecht et al. (2005) was unable to identify vernalisation mediating genes *FLC* and *FRI*. While photoperiodism response gene *CO* and *CO-like* genes *(COL)* showed no functional floral induction role in Medicago and pea (Weller et al., 2009, Putterill et al., 2013). This suggests that within the legume family there exists a novel floral regulatory pathway to arabidopsis, which potentially has been substituted by the expansion and sub functionalisation within the *FT* and *TFL1* genes.

1.7 Aims and Scope of this study

Crop domestication has been a major cornerstone in the development of human civilisation; therefore, understanding this process from an evolutionary perspective has been a topic of great scientific interest. As with most biological processes on an ecosystem scale, domestication is a highly complicated and dynamic process (Meyer and Purugganan, 2013).

This study will investigate aspects of the underlying genetic control of domestication in the crop and model legume pea (*Pisum sativum* L.). With pea being among one of the first crops to be domesticated, as well as being a highly important food crop, this study will have interest for evolutionary biologists and plant breeders. Current understanding has been limited due to the absence of a fully sequenced genome.

This study focuses on a biparental cross between modern cultivar *P. sativum var. sativum* line NGB5839 and wild *P. sativum var. humile* line JI1794*,* a representative of the "northern humile" lineage which is considered a major contributor to the modern day domesticated pea. It will explore

these questions by QTL mapping of domestication related traits and other morphological variations following the initial development of a high-density linkage map.

The genetic control of certain traits (primarily flowering time and seed dormancy) will then be examined in more detail using segregating populations developed from advanced generation progeny of the original NGB5839 x JI1794 F2. This study will provide a window into the evolutionary history of the cultivated pea as well as providing a platform for further research. It may also have commercial applications in plant breeding programs through marker assisted selection.

Chapter 3 will describe the development of the high-density pea linkage map which was developed for a population of RILs for the NGB5839 x JI1794 cross. The synteny of this map will be compared with other legume genomes and the feasibility of these relationships for application of a candidate gene approach will be discussed.

Chapter 4 will explore the "genetic architecture" of a range of traits, including many that are considered part of the "domestication syndrome", using QTL analysis. This will provide the foundation for later chapters to further investigate specific QTLs relating to seed dormancy and flowering time.

Chapter 5 will focus on the physical seed dormancy QTL on LGII (*Perm2*), and will further genetically dissecting this QTL region using advanced generation segregating populations. It will also investigate the potential effects of pigmentation loss and reduced testa thickness observed in the domesticated line.

Chapter 6 will provide further insight into the genetic control regulating physical seed dormancy. This chapter will focus on seed coat structural components, testa thickness and seed coat roughness mapped to LGVI and its association with seed coat permeability.

Chapter 7 will attempt to identify causative genes relating to long-day flowering time QTLs, using candidate gene approach and RNA expression. This will focus on QTLs on LGII (*DTF2*) and LGV (*DTF5*) which were previously not identified in short-day conditions, undertaken in a previous study (Weller et al., 2012).

Chapter 8 will summarise findings from this thesis, discussing the implications that crop domestication has had on the pea's genetic architecture as highlighted in chapter 4. This chapter will also discuss gaps and limitations within this field of study and highlight future prospective lines of research.

Chapter 2: General materials and methods

This chapter provides an overview of the materials and methods used in this project. Any additions or alterations that may occur within specific experiments will be described in the chapters accordingly.

2.1 Plant material

2.1.1 Recombinant Inbred Line population

An F8+ Recombinant Inbred line (RIL) population consisting of 138 lines was developed at the University of Tasmania by Jim Weller and Jackie Vander Schoor. This was derived from a wide cross between modern cultivar NGB5839 (*Pisum sativum var sativum*) and representative wild line JI1794 (*P. sativum var humile)*, which has been proposed as a primitive but major contributor to domesticated gene pool (Hoey et al., 1996, Zohary et al., 2012).

2.1.2 Advanced generation segregating populations

The advanced generation segregating populations, were developed by preferentially selecting lines from the earlier F2 RIL parental population. This consisted of 92 individuals which had previously genotyped using 88 markers across its seven linkage groups (17 + 3 + 32 + 0 + 14 + 19 + 3 = 88). When necessary the F3 advanced generation segregating population was further advanced to either narrow interval or increase population size. A detailed list of populations used in this study are detailed in Table 2.1.

Table 2. 1: Details of populations developed in this study. For details regarding genetic background refer to respective chapter.

Population name	Ancestor	Generation	Chapter	Population size
RIL	NGB5839 x JI1794	F8 to F9	3 and 4	138
Hr-Bc6		F2	4	16
TT2-3A	d/1/13	F3	5	17
TT2-3B	b/1/24	F3	5	7
TT2b-4A	d/1/13/6	F4	5	18
TT2c-4A	d/1/13/12	F4	5	17
TT2d-4A	d/1/13/15	F4	5	17
TT2b-4B	b/1/24/5	F4	5 and 7	51
TT2-5B	b/1/24/5/10	F5	5	
QTL6	d/1/27/9 d/1/27/10 d/1/27/11	F4	6	360
TT6-3A	e/1/17	F3	6	30
TT6-6A	d/1/27/11/52/1	F6	6	12
TT6a-6M	d/1/27/11/52/2	F6	6	49
TT6b-6M	d/1/27/11/52/10	F6	6	36
GTYa-5A	d/1/29/10/52	F5	6	44
GTYa-5B	d/1/27/9/51	F5	6	46
GTYb-5A	d/1/29/9/14	F5	6	27
GTYb-5B	d/1/29/9/24	F5	6	26
GTY-5AH	d/1/29/10/56	F5	6	16
GTY-5HB	d/1/29/11/11	F5	6	16
GTY-3AH	d/1/20	F3	6	16
GTY-6BH	d/1/29/10/52/42	F6	6	16
GTY-6BB	d/1/29/10/52/8	F6	6	16
FT2+5-3A	e/1/27	F3	6	16
TT2a-4B	b/1/21/4	F4	7	34
FT5a_4A	e/1/13/3	F4	7	19
FT5b_4A	e/1/13/4	F4	7	26
FT5_4M	e/1/27/4	F4	7	24

2.1.3 Additional plant material

Details of any additional plant material used in this study are described in relevant chapters.

2.2 Growth conditions.

Prior to sowing, seeds were scarified and treated with a fungicide Thiram. Each seed was sown in a 14cm pot prepared with a 1:1 gravel, vermiculite mixture, with an additional 3cm covering of sterilised potting mix which included controlled release fertiliser (CRF).

All plants were kept in controlled environment at the University of Tasmania and grown under long-day (LD) photoperiod conditions (16 hours light, 8 hours dark). Plants were watered at regular intervals dependent on growth cycle and growing season. Weekly nutrient, pesticide fungicide treatments were applied.

2.3 Phenotyping

Details of plant measurements as described in specific chapters

2.4 DNA and RNA extraction and quantification

2.4.1 DNA extraction

Fresh tissue samples were immediately frozen in liquid nitrogen and transferred to storage -71 freezers. Samples were ground using carbide bead and Mechanical cell lysis machine Qiagen TissueLyserII. 500 µl of 2x extraction buffer (100 mM Tris-HCl, 1.4M NaCl, 20 mM EDTA, 2% w/v CTAB, 20 mM 2-β-mercaptoethanol, pH 8 with HCl) is added to each sample and incubated at 60oC for 15 minutes with gentle agitation every 5 minutes. Solvent extraction was performed twice using 500 µl of chloroform-isoamyl alcohol (24:1) solution to purify samples. DNA is precipitated by adding 1ml of extraction buffer (50 mM Tris-HCl, 10mM EDTA, 1% w/v CTAB, pH 8 with HCl) and leaving for 15 mins before centrifuging for 10 minutes at 14000rpm. Solution is removed leaving pellet which is resuspended in 300 µl of a 1.5 NaCl solution containing 1 µl RNase A (25 mg/mL), this is optimised by incubating samples at 50oC with periodic agitation. Genomic DNA is precipitated for 15 minutes by adding 600 µl of 95% chilled ethanol, this is then centrifuged for 10 minutes at 140000rpm and ethanol is removed. DNA pellet is washed in 200 µl of 70% chilled ethanol and centrifuged for a further 5 minutes at 14000rpm to assist in the removal of ethanol. Any remaining residual ethanol is air dried before DNA is suspended in sterile Milli-Q water.

2.4.2 RNA extraction

Fresh tissue samples were immediately frozen in liquid nitrogen and transferred to storage -71 freezers. During sample preparation leaf tissues were ground using pestle and mortar and apex tissue using carbide beads and mechanical homogeniser (Qiagen TissueLyserII). RNA was extracted using the Promega SV Total RNA Isolation System (Promega, Madison, WI) according to the manufacturer's instructions. 1µg of total RNA was used to synthesise cDNA with Tetro Reverse Transcriptase (Bioline, London, UK) in a final volume of 20 µl following manufacturer's protocol. Negative control without reverse transcriptase (RT-) was included for all samples to check genomic DNA contamination. cDNA obtained was diluted five times for its final use.

2.5 Quantification and standardising of DNA and RNA Concentrations

DNA and RNA samples were quantified using a Nanodrop 8000 spectrophotometer (Thermo Fisher Scientific, USA) according to manufacturer's instructions. Concentration was determined from the average of three 3 readings. Using sterile Milli-Q water samples were standardised to 50 $ng\mu l^{-1}$ concentration.

2.6 Primer design

Primers were designed using the internet-based tool Primer3 version 4.0.0 (http://bioinfo.ut.ee/primer3/) and optimised according to criteria described in Table 2.2 Details of primers used are described in relevant chapters.

Table 2. 2: Details of primer design optimisation

Criteria	Range	Optimised to
Primer size (base pairs)	18 to 22	20
Primer Tm (°C)	55 to 62	60
Primer GC %	40 to 60	50
CG clamp	0 to 2	2
Product size	Dependent on function, see section 2.9	

2.7 Polymerase Chain Reactions (PCR)

2.7.1 Standard PCR

Standard PCR products were prepared to a final volume of 50 µl containing 5µl of DNA template, 10µl of 5 x buffer, 2µl of 50 mM $MgCl_2$, 1 µl of 10 mM dNTPs, 1µl each of forward and reverse primers, 0.1µl of MangoTaq™ DNA polymerase (Bioline, Australia) and 29.9µl of autoclaved Milli-Q water. Reactions were performed using a thermal cycler to the following conditions: 95°C for 5 minutes, 40 cycles [95°C for 45 seconds, annealing temperature (Tm; 50-62°C) for 45 seconds], 72°C for 2 minutes, 72°C for 10 minutes, HRM (temperature increasing with 0.1°C increments from 60-90°C, or from product melt temperature -5°C to +5°C).

Electrophoresis machine was used to test for amplification and contamination of PCR product by comparing product size against an appropriate ladder.

2.7.2 Real time polymerase chain reaction (qRT-PCR)

Relative gene expression using qRT-PCR were conducted on a Rotor-Gene 3000 Real-time Thermal Cycler with Rotor-Gene 6 Version 6.1 (Corbett Research, Australia). Reactions were prepared using either a Corbett Robotics CAS-1200TM pipetting robot (Corbett Research, Australia) or the PIRO Pipetting Robot (Lindauer DORNIER GmbH, Germany) with the software provided by supplier. Each reaction totalled volume of 10µL which consisted of 2 µl cDNA template, 5 µl 2X SensiFAST SYBR No-ROX mix (Bioline, Australia), 0.4 µM of each primer and autoclaved Milli-Q water.

To confirm samples were free of contamination controls containing water instead of cDNA were included. In addition, for each cDNA sample, ACTIN was run on the reverse transcriptase negative

control (RT-). Reactions were run for 50 cycles, and all samples were run in duplicate for greater accuracy.

Standard curves were performed to each target gene and within each qRT-PCR run. These were generated using a 10-fold serial dilution ranging from 10^{-1} to 10^{-6} ng/µl. Standard curve regression was considered acceptable if the R^2 value was equal to or higher than 0.99. Expression levels for each sample was calculated using the average of three technical replicates with each replicate standardised to *ACTIN* reference gene, in accordance to Hecht et al. (2011).

2.7.3 Visualisation of DNA

Visualisation of DNA fragments was performed using a BIO-RAD Molecular Imager® Gel Doc™ XR System. DNA fragments were separated using an electrophoresis on a 1.5–2% agarose gel in TAE buffer (40mM Tris Acetate and 1mM EDTA) containing GoldView™ Nucleic Acid Stain (Acridine orange; SBS Genetech Co., Ltd, Beijing, China). An appropriate DNA ladder was used to confirm product size.

2.7.4 PCR product purifying

PCR products were purified using Promega Wizard® SV Gel and PCR Clean-Up System (Promega, USA) and suspended in sterile Milli-Q water.

2.8 Sequence analysis

Purified PCR products were sequenced externally by Macrogen Inc. (Seoul, Korea). Sequences were analysed and annotated using Geneious software version 8 (Kearse et al., 2012).

2.9 Molecular marker design and genotyping

Molecular markers were designed only when no pre-existing markers were unavailable. Details of markers used in this thesis are described in relevant chapters.

2.9.1 Developing and scoring size markers

Size markers were developed when differences in amplified DNA product exceeded 15bp. If necessary new primers were developed to reduce the amplified product so that the corresponding deletion region was ≥10% of product size. Markers were scored against parental and heterozygous controls using standard PCR and visualisation protocols described in section 2.6.1 and 2.6.3.

PCR-based size markers were scored in segregating populations by standard PCR and visualisation of PCR products. Size differences identified through sequencing that were too small for PCR-based size markers were converted into HRM markers, which is described below.

2.9.2 Developing and scoring HRM markers

Primers were designed to flank and amplify small regions containing INDEL or SNP. Marker efficacy was evaluated by scoring marker across a segregating population, which in turn were compared against both parental and artificial heterozygous control samples. Each sample comprised of three technical replicates with each reaction prepared using either a Corbett Robotics CAS-1200TM pipetting robot (Corbett Research, Australia) or a PIRO Pipetting Robot (Lindauer DORNIER GmbH, Germany) with the software provided by supplier. Each reaction consisted of a 2µL DNA template, 1µL of each primer, 7.5 µl SensiFAST HRM Mix (from SensiFAST^{TM} HRM Kit, Bioline), and sterile milli-Q water to complete 15 µl.

HRM reactions were performed using a Rotorgene Q HRM machine (Qiagen) to the following conditions: 95°C for 5 minutes, 50 cycles [95°C for 10 seconds, annealing temperature (Tm; 50-60°C) for 30 seconds], 95°C for 5 minutes, 50°C for 5 minutes, HRM (temperature increasing with 0.1°C increments from 60-90°C, or from product melt temperature -5°C to +5°C).

HRM results were scored using Rotor-Gene® and ScreenClust HRM® Software (Qiagen).

2.9.3 Scoring CAPs markers

In this study no CAPs markers were developed however when necessary existing CAPs markers were used. Enzyme digests were performed according to manufacturer's instructions (New England Biolabs, Inc., Ipswich, MA) and were visualised using standard PCR protocols described in section 2.6.1 and 2.6.3.

2.10 Software and statistical analysis

Sequence editing and annotating was performed using Geneious 9.1.2 with sequence alignments conducted using MUSCLE (Edgar, 2004) plugin. Genetic maps were constructed using JoinMAP 4 and visualised using MapChart and QTL analysis performed using MapQTL 6.0. All statistical analysis was conducted using R studio. Comparative phenotype genotype association were conducted using Tukeys pairwise analysis with significance threshold set at ≤ 0.05. Standard error was used to show variation between technical replicates. When applicable figures were edited using Adobe illustrator. Details of all software used is described in Table 2.3.

Table 2. 3: Details of software and statistical programs used

Software	Version	Developer/Website	Reference
R studio	2.7.1	www.rstudio.com	-
Geneious	9.1.2	www.geneious.com	(Kearse et al., 2012)
JoinMap	4	www.kyazma.nl	(Stam, 1993)
MapQTL	6		(Van Ooijen, 2009)
MapChart	2.32	hwww.wur.nl/en/show/Mapchart.htm	(Voorrips, 2002)
Adobe illustrator		Adobe	-

2.11 Sequence resources

Online resources used for identifying gene homologs and aid in primer design and are listed in Table 2.4

Table 2. 4: Details of online resources used for sequence information

Resource type	Species	Resource	Version	Reference
Sequenced genome	Medicago	-	4.0	(Tang et al., 2014)
Sequenced genome	*Phaseolus vulgaris*	-	1.0	(Vlasova et al., 2016)
Sequenced genome	*Cicer arietinum*	-	2.0	(Parween et al., 2015)
Sequenced genome	*Trifolium pratense*	-	1.0	(Jan et al., 2014)
Sequenced genome	*Lens culinaris*	*Knowpulse*	1.2	*http://knowpulse2.usask.ca/portal*
Transcriptome	Pisum	-	-	(Tayeh et al., 2015) http://bios.dijon.inra.fr/FATAL/cgi/PsUniLowCopy.cgi
Genomic	Medicago	Phytozome	4.0	https://phytozome.jgi.doe.gov

Chapter 3: Development of a High-Density Pea Linkage Map

3.1 Introduction

The pea plant has been the model system from which classic genetics was founded (Knight, 1799, Mendel, 1866) and today, with its rich resource of mutant lines, it remains an important model for studying plant genetics. Molecular research has accelerated with the advancement in high-throughput genotyping technologies, allowing large scale sequencing programs more accessible. To date, 25 fully sequenced and annotated plant genomes are available (Geleta and Ortiz, 2016) however, despite its historic and economic significance no fully sequenced genome is currently available for pea.

The advancement of the pea genetic map has been a gradual process. Originating from morphological (Blixt, 1972) and low-resolution molecular marker maps (Weeden and Marx, 1987, Weeden and Wolko, 1990), the first consensus map to include peas seven recognisable linkage groups was completed by Weeden et al. (1998). Since then more detailed linkage maps have been developed, typically comprising between 1000 – 2000 markers (Duarte et al., 2014, Sindhu et al., 2014, Sudheesh et al., 2015, Ma et al., 2017) with the most comprehensive transcriptome map, consisting of around 13.2K markers (Tayeh et al., 2015), nonetheless to-date no fully sequenced genome for pea is available. Its protracted advancement has in part been caused by its extensive genome size, which has been calculated at ~4.5Gb (Greilhuber and Ebert, 1994, Macas et al., 2007), making analyses of its structure and organisation problematic (Ellis and Poyser, 2002). As shown in Figure 3.1, the pea genome is much larger than most other legume species but is comparable with other species in the *Fabeae* tribe. This includes well known species *Lens culinaris* ~4Gb, *Vicia* ~5.4Gb (including *Vicia faba* estimated ~13Gb (Cooper et al., 2017) and *Lathyrus* ~7.2Gb (Bennett MD, 2012). Its expansion has predominately resulted from the accumulation of repetitive DNA mobile elements, which occurred with the divergence of the *Fabeae* tribe (Macas et al., 2007). This is consistent with the closest relative to the extinct common ancestor of the *Fabeae* tribe, *Vavilovia* (Smýkal et al., 2011, Schaefer et al., 2012, Ochatt et al., 2016) which has a relatively intermediate genome size of around 2.32Gb (Mikić et al., 2013), highlighting its progressive genome expansion.

a)

Species	Genome size	Unmapped markers	Karyotype	Version	Reference
Vicia faba	13Gb	-	2n = 12	-	(Cooper et al., 2017)
Lathyrus spp	7.2Gb	-	2n = 14	-	(Bennett MD, 2012)
Pisum sativum*	4.5Gb	-	2n = 14	-	(Greilhuber and Ebert, 1994, Macas et al., 2007)
Lens culinaris*	4Gb	-	2n = 14	1.2	(Knowpulse, 2018)
Vavilovia	2.32Gb	-	2n = 14	-	(Mikić et al., 2013)
Arachis ipaensis	1.4Gb	11.1Mb	*2n = 20*	1.1	*(Bertioli et al., 2016, Chen et al., 2016)*
Arachis duranensis	1.1Gb	43.23Mb		1.1	
Glycine max	955.37Mb	29.31Mb	2n = 40	2.1	(Shimomura et al., 2015)
Vigna angularis	612Mb	94.01Mb	2n = 22	1.1	(Kang et al., 2015)
Cajunus cajan	558.38Mb	345.32Mb	2n = 22	1.0	(Varshney et al., 2011)
Lupinus angustifolius	605.05Mb	138.78Mb	2n = 40	1.0	(Hane et al., 2017)
Vigna radiata	463.09Mb	129.78Mb	2n = 22	1.0	(Kang et al., 2014)
Cicer arietinum*	482.62Mb	183.52Mb	2n = 16	2.0	(Parween et al., 2015)
Phaseolus vulgaris	472.45Mb	6.26Mb	2n = 22	1.0	(Vlasova et al., 2016)
Lotus japonicus	394.46Mb	-	2n = 12	3.0	(Kazakoff et al., 2012)
Medicago truncatula*	389.89Mb	28.33Mb	2n = 16	4.0	(Tang et al., 2014)
Trifolium pratense*	338.78Mb	156.81Mb	2n = 14	1.0	(Jan et al., 2014, Ištvánek et al., 2017)

b)

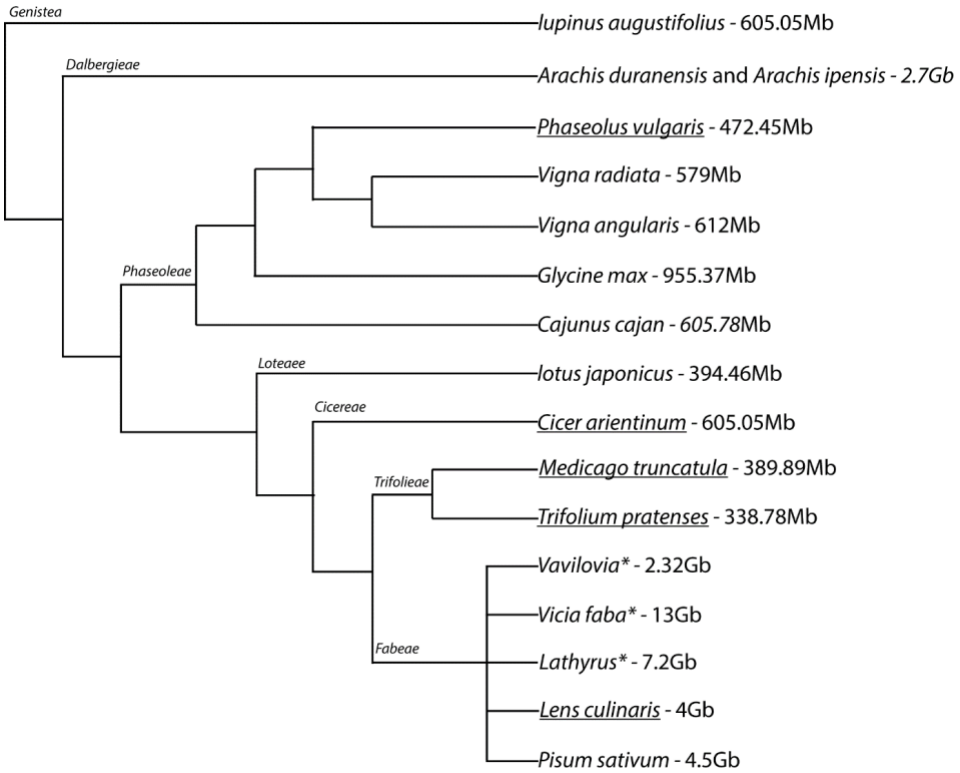


*Figure 3. 1: Details of size and taxonomy of sequenced genomes within the legume family. a) Table detailing genome sizes b) schematic representation of the phylogeny based from (Bruneau et al., 2013, Azani et al., 2017) and genome sizes references given in text. * signifies genomes that have not been sequenced this includes genuses Vavilovia Lathyrus where sizes are calculated by average C-values of all species within genus (Bennett MD, 2012) and Vicia faba estimated using transcriptome data. Species underlined will be used for comparative analysis with Pea*

In the absence of a fully sequenced genome, functional genetic studies have relied on synteny between closely related species. The usefulness of a reference genome is dependent on confidence to its genome assembly and syntenic relationship (Tayeh et al., 2015, Lee et al., 2017). In pea, genetic studies have primarily relied on *M. truncatula* due to its well-established genome and close

phylogenetic affiliations (Kaló et al., 2004, Tang et al., 2014). In general pea and *M. truncatula* have a close syntenic relationship, however regions of poor synteny occur as seen on PsLGVI between the Mt2 and Mt6 translocation event (Kaló et al., 2004, Tayeh et al., 2015). Consequently, the genomic structure in these regions are poorly understood which can constrain functional genetic studies.

The recent release of the *T. pratensis* (Jan et al., 2014) and *L. culinaris* (Knowpulse, 2018) sequenced genomes potentially offers better models to infer pea genomic structure due to their close phylogenetic affiliations. *T. pratensis* belongs to the same tribe as *M. truncatula* (*Trifolium*), but like pea has one less chromosome, therefore indicating a high level of synteny with pea. To date, *L. culinaris* is the most closely related fully sequenced genome available for pea, sharing similar genomic size and structure and therefore implicating this as its most syntenic available genome. To our knowledge, there has been no published comparative analyses between pea and *T. pratensis,* whereas for *L. culinaris* only low resolution analyses have been made (Weeden et al., 1992, Lee et al., 2017) but none using its fully sequenced genome.

Recent pea linkage maps have primarily been developed using *Diversity Arrays Technology* (DaRTseq) and have comprised between 1000-2000 markers (Duarte et al., 2014, Sindhu et al., 2014, Sudheesh et al., 2015, Ma et al., 2017). While these maps are adequate for QTL analysis they become limited for determination of microsynteny, which is important in further fine mapping studies. The capacity of Next Generation Sequencing technologies to generate large number of markers has dramatically increased the potential map resolution, an approach that was recently used for the 13.2K high-density pea consensus map (Tayeh et al., 2015). This was developed by integrating 12 recombinant inbred populations derived from a diverse range of pea parental lines. This map has become the best current reference for inferring pea genomic structure and will be a valuable resource for candidate gene approaches later in this thesis.

As outlined in chapter 1, the overall purpose of this thesis is to investigate the genetic basis for pea domestication by conducting a QTL analysis in a wild x cultivated biparental population, for which a linkage map is not yet available. This chapter aims to construct a high-confidence, high-density linkage map for a Recombinant-Inbred Line (RIL) population derived from the cross NGB5839 (a derivative of cv. Torsdag) and the wild line JI1794, a representative of the northern race of *Pisum elatius* (*var humile*) that is thought to be the most likely ancestor of domesticated pea. This population has genotype information for over 6000 sequence-tagged DArTseq markers which are available for map construction. Comparative analysis will be performed with five closely related species; *L. culinaris*, *T. pratense, M. truncatula, C. arietinum* and *P. vulgaris*. The resulting map will be used for QTL analysis

in Chapter 4 and synteny information will be used to guide the candidate gene approach and development of targeted markers in subsequent chapters.

3.2 Methods

3.2.1 Population development, DNA extraction and genotyping

A F_8+ RIL population comprising of 138 individuals was developed from a biparental cross between representative wild (*P. sativum var humile*) and domesticated (*P. sativum var sativum*) lines JI1794 and NGB5839 respectively. This development work occurred at the Hobart University, Tasmania by Jim Weller and Jackie Vander Schoor from an F_2 population described in Weller et al. (2012). Prior to the initiation of this project, genomic DNA was collected from parental lines and all 138 individuals at the F_8+ generation and extracted using protocol described in chapter 2.

A total of 6238 DArT were generated and genotyped using the DArT-Seq high-throughput method (http://www.diversityarrays.com/), by Diversity Array Technology Pty. Ltd., Canberra Australian Capital Territory. In addition to DArTseq markers, 24 other gene-based anchor markers (Table 3.1) were developed, genotyped and incorporated to the map to facilitate characterization of synteny in certain specific genomic regions.

*Table 3. 1: Details of anchor markers mapped in pea, *signifies scaffold in Medicago*

Marker	Primer	Primer sequence	Type of Marker	Gene ID	Expected pea LG
Mendel A	N/A	N/A	Phenotype	Medtr1g072320	2
AGOI	AGO1-HRM-F2	TTACTCCCATGTCATCCTTGG	HRM	Medtr6g477980	6
	AGO1-HRM-R2	CAAGCATTAAAGAACCAGCAAG			
APRL	PsAPRL-3F	TGGGATGCTTCCTATTGGTT	HRM	Medtr6g029240	6
	PsAPRL-3R	TGAACATGGTCTGAAAATCTCAC			
BFT	BFT-HRM-17F	GGCCAATTTTGCTGATGACT	HRM	*	6
	BFT-HRM-17R	TTTGACCACACTTGGTTCAAC			
BRC2	BRC2-1794-F	AACCAGCTTAATTTTCTTCTTTG	HRM	Medtr6g017055	6
	BRC2-1794-R	CACTCAAAATTATGGAATTTCAGG			
CABB	CABB-F2	AGGATCTTCTTGCCTGATGG	HRM	Medtr6g060175	6
	CABB-R3	CTTGCTTAGACCAAAAGGATCA			
COP13	Ps-COP13-4F	ATAAAAGTTGATATGGGAGAAAGA	HRM	Medtr6g023350	6
	Ps-COP13-4R	CCAATGCAGGCACTCATA			
CYTB7	CYTB7-1F	TCAAGGAGGCTCTGAATCGG	Size	Medtr1g011880	2
	CYTB7-1R	CCTGAATGGTGTTTGCATTGC			
DUF	DUF-F1	GTGGCAAGCTCATCCAAAAT	HRM	Medtr4g083440	7
	DUF-R1	CTACCGCCCAAGATTGTTTC			
ERMP	ERMP-3F	CATCAGATTGGGTGTCGTCC	Size	Medtr1g013660	2
	ERMP-3R	AGACCACCACCTGAGTATTCC			
FTa1	FTa1-1794-F	GGACGTGAGCAAAACGACAT	HRM	Medtr7g084970	5
	FTa1-1794-R	TTGAGTAGTACCAGCACACACT			
FTa3	FTa3-F1	TTGTTCTTGGAGCTGTAATTGG	HRM	Medtr6g033040	6
	FTa3-R1	CCTCAAATTTGGGTTACTAGGG			
FTc	FTc-1794-F	CATTGGGATGTTAAAATGGTG	HRM	Medtr7g085040	5
	FTc-1794-R	TGGGAAAGAGTTGCAAGATG			
FULa	PM2-HRM-F	AACCTAGTAGCTCTCACCGTAA	HRM	Medtr2g461760	6
	PM2-HRM-R	TTATATTATGGTGTTTGATTGATGA			
Ga20ox	GA20ox-2F	GACCAACTTTTTAAGAAAAGCA	HRM	Medtr6g464620	6
	GA20ox-2R	TCTCCCATTTGAAAGAGCCTA			
HR	ELF3-HR-F	ACTAACACTTTATTGGCAAGTG	HRM	Medtr3g103970	3
	ELF3-HR-R	GCGGAAAGTATCGTCATTTTG			
JMJ	JMJ-3F	CTAGAGTGAAGTGTAATTGTAAG	HRM	Medtr1g078070	2
	JMJ-3R	TGCCAGAATAAGGAAAATGGAG			
LE	PsLe-HRM-1F	TGTCGTGCAATATGATGAAACC	HRM	Medtr2g102570	3
	PsLe-HRM-1R	CGGCCCATTTGATATCTTCC			
LF	LF-CAM-F	GGTCCCTCTTTACCCTGGTATT	HRM	Medtr1g060190	2
	LF-CAM-R	TGATCTGCAGGAAAACAATAAA			
MLO1	MLO1-7F	TGGCTCTTAGGCATGGATTT	HRM	Medtr6g033330	6
	MLO1-3R	TTGTGCATCATGTCCTGGAG			
MTIC153	MTIC153-HRM-1F	TGCAACAAAAGAGGTATGAACTG	HRM	*	6
	MTIC153-1R	TGGGTCGGTGAATTTTCTGT			
NT6083	NT6083-HRM-1F	CGTGTTTTTCTGAGTTGACTTCC	HRM	Medtr6g034195	6
	NT6083-HRM-1R	TGTATACAGGGCAAACTCCTTG			
RNAhel	RNAhel-F	GGGTTTGGTAGGTTTGGTAGAGG	HRM	Medtr6g056080	6
	RNAhel-HRM-R	GCATGTGCTATTTTTCTTCACTC			
TMP	TMP-1F	CACCCACAAATCCCTCTTCC	Size	Medtr1g017450	2
	TMP-1R	AACAGCCCATGATTTAGCGG			

3.2.2 Constructing genetic linkage Map

A linkage map was developed using JoinMap version 4.0. Markers were assigned into the 7 Linkage Groups (LG) by adjusting the independence *Logarithm of Odds* (LOD) significance threshold. For each linkage group, markers were ordered using the maximum likelihood model, with *Kosambi* mapping function used to convert recombinations into mapping distance.

A second marker order assessment was implemented using the regression mapping algorithm using the same *Kosambi* mapping function. Marker order from the maximum likelihood and regression mapping models were compared using the *MapChart* software (Function "show homologs"). Segregation distortion across map was calculated using JoinMap (P = < 0.01).

3.2.3 Evaluating syntenic relationship between *P. sativum* and reference genomes

Synteny of the newly-generated NGB5839 x JI1794 map were evaluated against five closely related legume species; four from the *galegoid* clade (P. sativum, *Trifolium pratense*, L. *culinaris, M. truncatula and C. arietinum)* and one species from the *phaseoloid* clade (*P. vulgaris*) (see Figure 3.1). Potential orthologous relationships between pea DArT markers and other legume gene markers were established by BLAST searches of markers used in the construction of the linkage map against individual reference genomes.

Mapping was performed using Geneious v9.1.2 software using custom sensitivity set to the following parameters: Map multiple best matches: none; Trim paired reads; Minimum support for structural variant discovery: 2 reads; Allow gaps set to a maximum of 5 per read and a maximum of 3; Word length: 6; Index word length 6; Maximum mismatch per read: 35%; and Maximum ambiguity: 5. Fine tuning: none.

Mapping was finalised by reviewing marker alignments to remove potential marker mapping duplication. If duplication occurred, only the marker in the syntenic position with respect to the reference genome was retained. If mapping position could not be resolved the marker was excluded from the analysis.

The syntenic relationship was visualised using a dot plot constructed using the R-studio program. For illustrative purposes, the sizes of each linkage group and chromosome was standardised by converting marker positions from calculated to relative position (0 to 100).

3.2.3.1 Identifying syntenic regions

Regions were considered syntenic if they consisted of >5 consecutive markers positioned with a distance between adjacent markers that were <10% of the total LG length. Once a syntenic block was established, subsequent markers were still required to adhere to the <10% rule but were permitted to have a maximum gap of three non syntenic markers. If a gap exceeded three markers, then a new syntenic block must be established, as previously described. In the comparative analysis syntenic regions were highlighted in colour.

Syntenic regions were categorised into three groups according to size of syntenic region, which was represented as proportion of collinearity in relation to LG size. This included large (≥ 50%), medium (<50 and ≥5%) or small (<15%).

3.3 Results

3.3.1 Development and genotyping

All DArT markers and other gene-based markers (6239 + 24 = 6263 markers) were scored across the 137 individuals.

3.3.1.1 Assessing marker quality

As an initial step in sorting and prioritizing markers for mapping, markers were assigned into four quality classes (from 1/best to 4/worst) based on their quality. Quality was assessed by four selection parameters (Table 3.2a) that were provided with the genotype dataset by Diversity Arrays Technology: i) Call rate, ii) Reproducibility, iii) Segregation distortion (PIC) and iv) Proportion of heterozygotes. Each marker was given a rank value for each selection parameter and an overall quality ranking was assigned based on the sum of the four rank values (Table 3.2b). Markers were then relabelled to indicate their rank quality, (e.g. rank 1 marker 3559539 represented as "3559539_1") for ease of tracking during the mapping process. Results from the marker quality assignment is summarised in Table 3.2c and detailed in Supplementary Figure 3.1.

Table 3. 2: Tables detailing a) DArT marker quality selection criterion b) assignment of marker quality c) results of markers quality.

a)

Rank Value	Call rate (%)	Reproducibility (%)	Segregation distortion	Proportion of Hets (%)
1	100	100	0.45 - 0.55	0 – <1.45
2	>98	>98.71	0.4 - 0.6	>1.45 – <3.63
3	>90	>98	0.3 – 0.7	>3.63 – <7.25
4	<90	<98	<0.3 - >0.7	>7.25

b)

Marker Quality	Total of all rank values	Unless
1	4-5	
2	6-7	Unless criterion = 3 then rank quality is 3
3	8-10	Unless criterion = 4 then rank quality is 4
4	11+	

c)

Rank Quality	Number of DArT markers	Number Anchor Markers
1	1275	8
2	1146	15
3	2161	1
4	1656	-
Total	6238	24

3.3.2 Constructing consensus map

3.3.2.1 Linkage map construction

As the number of molecular markers was too large for efficient map construction as a single computational task, an initial *Skeleton Map* was developed. This was constructed from a subset of markers using the marker "binning" process. This was applied using the SimpleMap "*before mapping"* function (Jighly et al., 2015) to reduce effective marker number for the initial mapping round, with minimal effect to marker coverage. The binning process involves the grouping of tightly linked markers according to a user-defined maximum recombination threshold (repulsion threshold), which was set at 4cM. From each bin (referred to as "bin markers") a single marker of highest rank quality was selected (referred to as "representative marker") to represent the co-segregating bin markers. Of the original 6263 markers, SimpleMap identified 2091 representative markers (see Supplementary Figure 3.2), further 853 lowest quality markers were also removed for mapping robustness, totalling 1238 representative markers. These were assembled into the seven linkage groups using JoinMap software, with the linkage threshold set at a minimum LOD 5 (which equates to probability P ≥ 0.000001), exceeding the traditional LOD 3 score threshold used in other studies (Nyholt, 2000).

Representative markers were ordered using an iterative mapping approach from highest quality to lowest quality markers, a similar technique utilised in (Butler et al., 2017). With each iteration, regions containing a significant conflict to marker mapping were identified using plausible fit function and markers removed based on stringent marker exclusion criterion (Table 3.3), with the exclusion of lesser quality markers taking precedence. This procedure was repeated for each linkage group until all conflicting regions were resolved.

*Table 3. 3: Marker exclusion criterion for conflicting regions during the construction of the skeleton map. ** P ⩾0.01*

Statistic	Deletion threshold
Genotype probability	Double inversion with ≥ ** significance
Fit and Stress	NN stress ≥ 2cM
Locus genotype frequency	Isolated markers with Segregation disequilibrium ≥ ** when placed in chronological marker order

As detailed in Table 3.4, a skeleton was developed using the maximum likelihood model comprising of 905 representative markers. For map robustness and to assess confidence in marker order the newly developed map was compared to a second skeleton map ordered using the same 905 markers with the regression mapping model. The comparison showed strong correlation with few minor rearrangements in localised areas (highlighted in red in Figure 3.2). The marker ordering from the maximum likelihood model was used for the remainder of this study because the marker inclusion/exclusions were based on these statistical calculations.

Table 3. 4: Table showing number of markers used in the construction of the skeleton map before and after marker exclusion.

	Number of markers		
	Initial	**Deleted**	**Final**
LGI	133	24	109
LGII	201	62	139
LGIII	249	90	159
LGIV	167	43	124
LGV	176	26	150
LGVI	142	34	108
LGVII	172	54	118
Total	1238	333	905

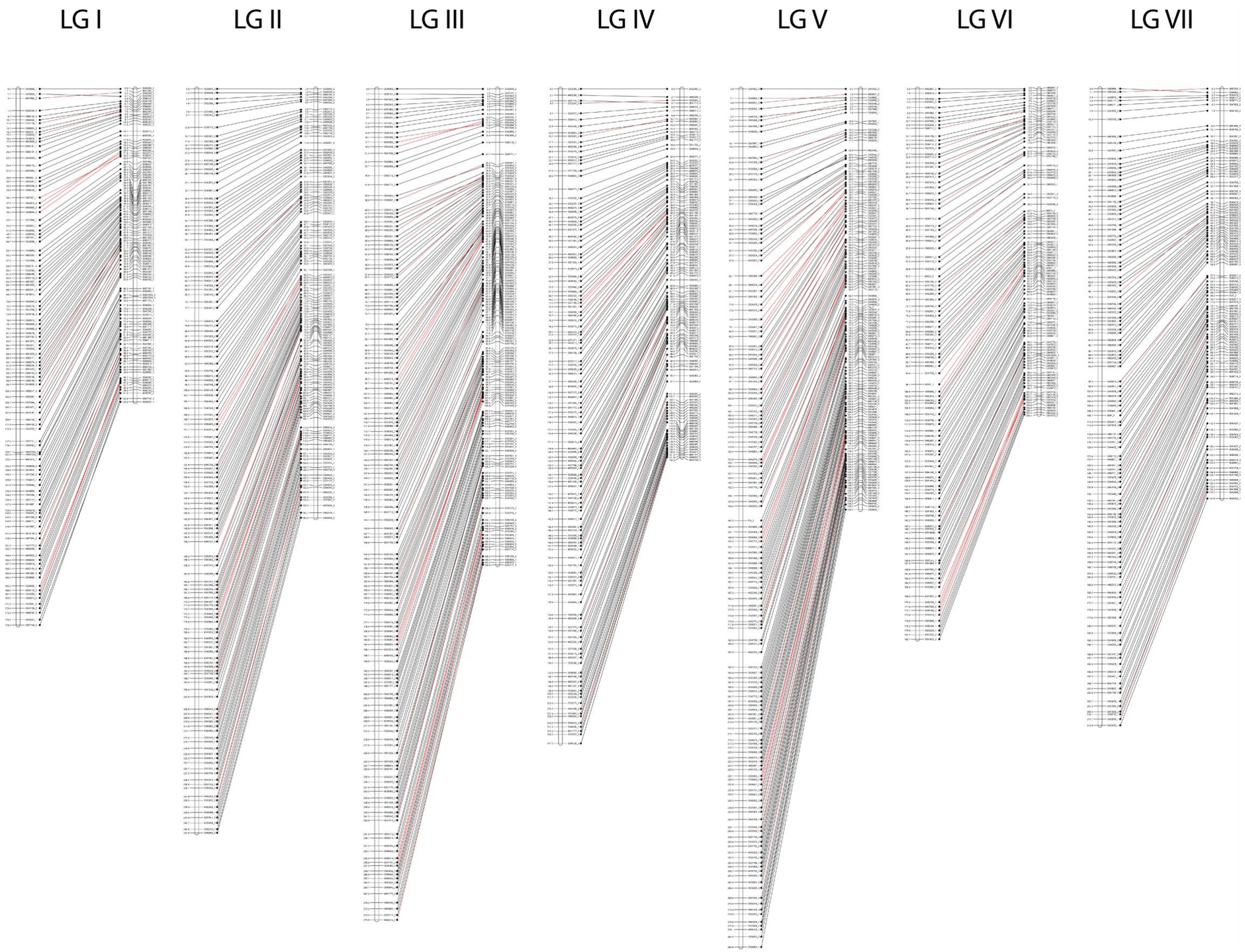


Figure 3. 2 Comparing marker order of the 905 representative markers across the seven linkage groups using two mapping methods i) Maximum likelihood Model (Left) and ii) Regression Mapping (right). Lines show homologous, black lines illustrate consistent marker order and red lines highlight variations. (High resolution version can be accessed in https://cloudstor.aarnet.edu.au/plus/s/E3WKRxNqWi5qXL3)

The linkage map was finalised by inserting the remaining markers from each bin back into the skeleton map. As shown in Figure 3.3 and detailed in Table 3.5, the final linkage map consisted of 4599 markers (4575 DArT and 24 anchor markers) with the number of markers per linkage group ranging between 525 - 857. The linkage map spanned a total of 1617cM, ranging from 182 - 279cM in length per linkage group. The average distance between adjacent markers was 0.35cM and ranged from 0.30 - 0. 52cM between the different linkage groups. There were no gaps between adjacent markers that exceeded 10cM and only four instances where gaps were greater than 5cM. The largest gap was 5.8cM on LGVII between markers 4661894_3 and 3563769_3. This was followed by 5.1cM on LGVII and on LGV between markers 3555709_3 and 3555057_3, and markers 3564306_3 and 3564306_3 respectively, (see supplementary Figure 3.3). The segregation distortion (P = < 0.01), estimated using initial skeleton map, averaged at 12.8%. LGVII had a significantly higher segregation distortion compared with the other linkage groups, of 44.1%.

Table 3. 5: Summary of the high high-density consensus map.

	LGI	LGII	LGIII	LGIV	LGV	LGVI	LGVII	Total
Markers in skeleton map	109	139	159	124	150	108	118	905
Markers added	416	508	700	578	408	512	593	3692
Combined Markers	525	647	857	702	558	620	711	4599
Size (cM)	182.31	249.86	279.49	217.34	288.53	182.70	216.73	1616.96
Density marker/cM)	2.88	2.59	3.07	3.94	2.43	3.39	3.28	3.08 (average)
Number of gaps (> 10cM)	0	0	0	0	0	0	0	0
Segregation distortion % (P = <10)	0	11.51	15.92	12.1	0	6.31	44.07	12.84 (average)

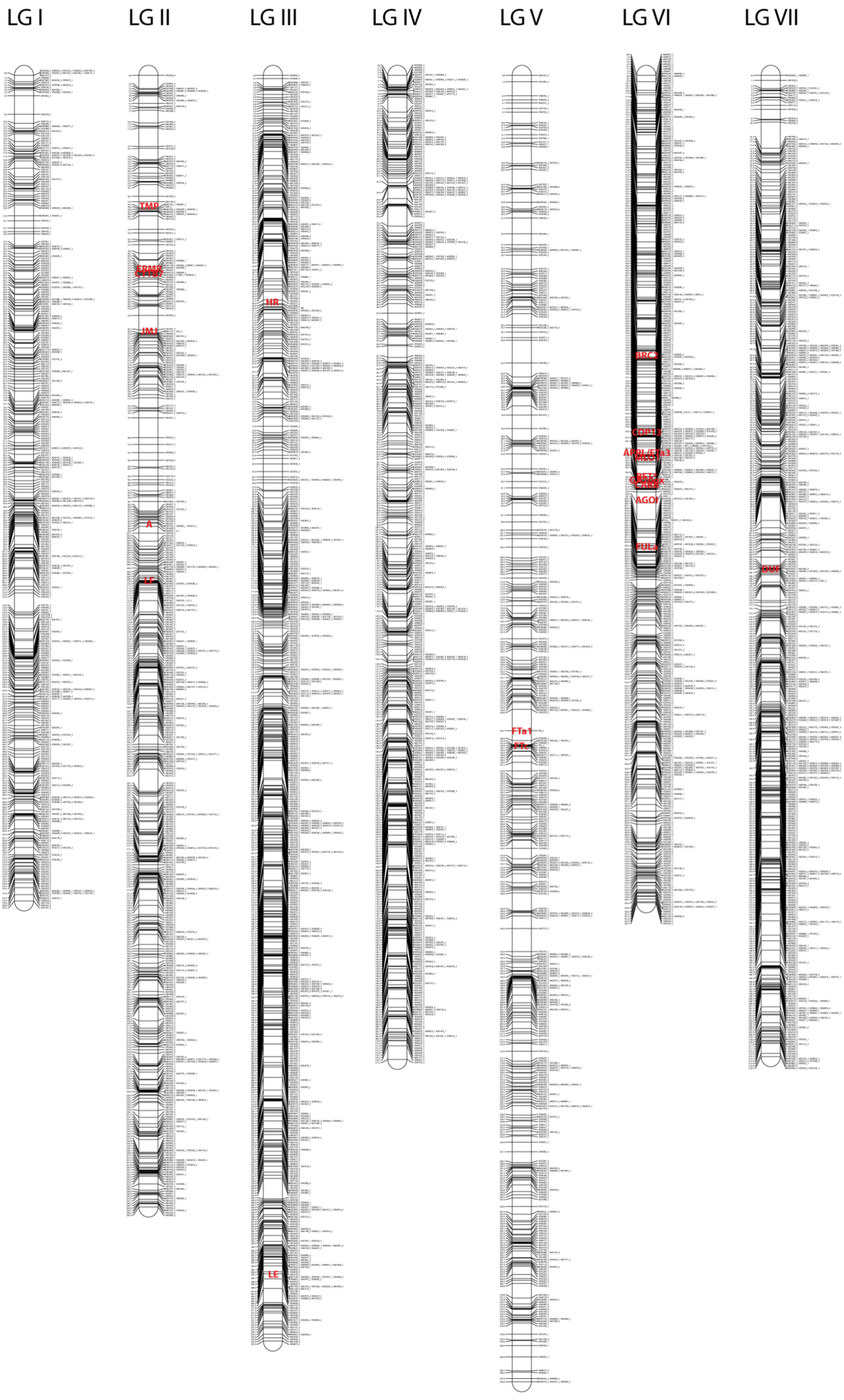


Figure 3. 3: Finalised NGB5839xJI1794 high density linkage map consisting of 4599 markers. Anchor markers are highlighted in red, BFT includes anchor markers MTIC153 and NT6083 which between them showed no recombination events. (High resolution version can be accessed in https://cloudstor.aarnet.edu.au/plus/s/E3WKRxNqWi5qXL3)*

3.3.3 Map comparisons

3.3.3.1 Comparisons with other pea maps

The structure of the new map was compared with six *P. sativum L.*, and one *P. fulvum* recently developed linkage maps, this includes the first map to give the seven pea linkage groups (Weeden et al., 1998). As shown in Table 3.6, the relative size per LG was relatively uniform, with outgroup species *P. fulvum* showing greatest variation from the consensus. Overall, LGI and LGVI were generally the smallest and LGIII and LGVII the largest. This was consistent with our map (Map 1) and previous cytogenetic maps (Ellis and Poyser, 2002), however LGV was larger than general consensus.

For each meiosis event this study identified an average of 4.6 crossovers per linkage group per meiosis event. This is comparable, albeit on the higher level, to the seven previous linkage maps, which when calculated ranged between 2.2 and 6.8 crossovers per linkage group (Table 3.6). It is likely that this study has overestimated the number of crossovers due to the inclusion of lower quality markers in the development of the map with genotyping errors causing false positive crossovers recordings, despite incorporating a marker quality exclusion criterion. This is supported by a cytogenetic analysis where 2 to 2.5 crossovers per linkage group was reported (Hall et al., 1997).

Table 3. 6: A comparison of the relative size per LG of the map developed in this study (map 1) with six previously constructed maps (maps 2-7). These were developed using accessions (1) NGB5839 x JI1794, (2) JI1794 x Slow, (3) VavD265 x Ballet, VavD265 x Cameor, Ballet x Cameor, Sommette x Cameor, Cerise x Cameor, China x Cameor, Kazar x Cameor Melrose x Cameor, Kazar x Melrose, Champagne x Terese, Baccara x PI180693 and JI296 x DP, (4) Kiflica x Aragorn, (5) Champagne x Terese, (6) Orb x CDC striker, Cameor x China, Alfetta x P651, CDC Bronco mutugenic line 1-2347-144 x CDC Meadow and Carerra x CDC Striker, (7) Kaspa × Yarrum, Kaspa × ps1771, (8) IFPI3260 x IFPI3251. Colour scale used to inform smallest (blue) to largest (white) LGs.

		Size (cM)									
map	Species	LG I	LG II	LG III	LG IV	LG V	LG VI	LG VII	Recombination events per linkage group	Total No. of markers	mapping function	Reference
1	*P. sativum L.*	182	250	280	217	289	183	217	4.6	4599	Kosambi	This study
2	*P. sativum L.*	97	130	130	103	117	85	100	2.2	850	-	(Weeden et al., 1998)
3	*P. sativum L.*	93	114	135	115	113	101	122	2.3	13204	Haldane	(Tayeh et al., 2015)
4	*P. sativum L.*	158	178	227	169	190	177	212	3.7	1683	Kosambi	(Ma et al., 2017)
5	*P. sativum L.*	147	218	203	169	156	142	220	3.6	2070	Haldane	(Duarte et al., 2014)
6	*P. sativum L.*	113	90	134	121	107	89	118	2.2	1536	Kosambi	(Sindhu et al., 2014)
7	*P. sativum L.*	337	356	411	353	233	309	388	6.8	2028	Kosambi	(Sudheesh et al., 2015)
8	*P. fulvum*	308	350	281	220	266	223	231	5.4	12058	Kosambi	(Barilli et al., 2018)

3.3.3.2 Comparisons with other legume genomes

Data describing the syntenic relationships between pea and five closely related legume species (Figure 3.1) are presented in Figures 3.4 to 3.8. This was achieved by comparing the orthologous positions of the 4599 markers used in the construction of our map, 30 markers were excluded due to overlap in marker position, therefore totalling 4569 markers.

As moderate to high detailed comparisons in pea have previously been conducted with *M. truncatula* and *C. arietinum* (Duarte et al., 2014, Tayeh et al., 2015, Ma et al., 2017) these were used to authenticate the establishment of our map. There have been no detailed comparisons made with recently released *T. pratense* and *L. culinaris* genomes and these were analysed to determine their suitability for functional genetic studies. *P. vulgaris* is the most distant related species compared in this study and will be used as an outgroup to investigate the effects of synteny with increased phylogenetic distance.

The total number of markers mapped to reference genomes ranged from 1847 markers in the recently developed *L. culinaris* genome to 3903 markers in *C. arietinum*. The percentage of mapped markers considered syntenic ranging between 30.7 in *P. vulgaris* to 72% *L. culinaris*. In all comparative analyses, apart from *L. culinaris*, pea LG (PsLG) VI had the lowest proportion of syntenic mapped markers. This corresponded to a region known to have a poor syntenic relationship with other legume species, such as *M. truncatula* (Kaló et al., 2004) and *C. arietinum* (Tayeh et al., 2015). This occurs around a major translocation event which interestingly was detected in *L. culinaris*. As expected, the number and sizes of syntenic blocks differed in the different comparisons (Table 3.7).

Table 3. 7: Macroscale synteny between pea linkage groups and corresponding model species chromosomes. Size of syntenic blocks categorised as L = Large (⩾50% Ps LG); M = medium (<50%, ⩾5% of Ps LG); S = Small (<5% of PsLG). Numbers indicate corresponding chromosome.

P. sativum	*L. culinaris*	*M. truncatula*	*T. pratense*	*C. arietinum*	*P. vulgaris*
LG I	1^{L}, 5^{S}	5^{L}	2^{M},4^{M},5^{S}	2^{L}, 8^{M}	4^{L}, 10^{M}
LG II	5^{L}, 1^{M}	1^{L}	1^{L}, 3^{S}, 6^{S}	4^{L}	5^{M}, 11^{M},9^{S}
LGIII	3^{L}, 2^{M}	3^{L}, 2^{M}	1^{S}, 2^{M}, 3^{M}, 7^{L}	5^{L}, 1^{M}, 4^{S}, 6^{S}	3^{L}, 7^{M}, 6^{M}, 11^{S}
LG IV	7^{L}	8^{L}, 4^{L}, 5^{S}	1^{S}, 2^{M}, 3^{S}, 4^{M}, 5^{M}, 6^{S}	7^{L}, 8^{L}	9^{L}, 2^{M}, 4^{S}
LG V	6^{L}	6^{L}	3^{S}, 5^{S}, 6^{L}	3^{L}, 2^{L}	11^{M}, 3^{M}, 5^{S}
LG VI	2^{L}	2^{L}, 6^{M}	2^{M}, 7^{S}	1^{L}, 2^{M}	6^{M}, 3^{M}, 8^{S}, 7^{S}, 10^{S}
LG VII	4^{L}	4^{L}, 8^{M}	1^{S}, 2^{M}, 3^{L}, 4^{M}	6^{L}	10^{M}, 1^{M}, 2^{S}, 4^{S}

3.3.3.2.1 Lens culinaris (Lentil)

Like pea, lentil also belongs to the *Fabeae* tribe, comprising of the same number of LGs and is the closest available sequenced genome. To date, comparative analysis between pea and lentil has been based on low resolution orthologous markers (Weeden and Marx, 1987, Weeden and Wolko, 1990) or indirect comparative analysis using *M. truncatula* to infer syntenic relationship (Sindhu et al., 2014). This will be the first high resolution comparison using its fully sequenced genome.

As, shown in Figure 3.4, strong collinearity of pea linkage groups PsLGI to LGVII was found with lentil linkage groups LcLG5, 1, 3, 7, 6, 2 and 4 along their entire lengths, which is consistent with previous reports (Sharpe et al., 2013, Sindhu et al., 2014). However, several small differences were noted, including small translocations on PsLGI, II and III, a small inversion on PsLGVII, and multiple inversions

on LGII. The translocations found at the top of PsLGIII corresponds to same translocation in *M. truncatula* chromosome (Mt)2. Two translocation events previously not reported were found at the top of PsLGII and V which corresponded to LcLG5 and 1 respectively.

		Hits in *L. culinaris*										
		Lc1	Lc2	Lc3	Lc4	Lc5	Lc6	Lc7	Total	mapped	unmapped	Syntenic (%)
Pisum	LGI	28 (22)	2	7	3	189 (148)	1	2	524	232	292	74
	LGII	243 (200)	0	6	0	47 (44)	1	2	638	297	341	83
	LGIII	2	63 (45)	258 (200)	2	11	8	2	854	346	508	72
	LGIV	1	2	3	2	17	3	233 (164)	701	261	440	64
	LGV	5	1	5	5	2	194 (123)	1	537	213	324	60
	LGVI	2	212 (159)	0	2	1	1	2	606	220	386	75
	LGVII	3	3	7	258 (198)	3	2	2	709	278	431	74
	Total	284	283	284	272	270	210	244	4569	1847	2722	-
	Syntenic (%)	78	72	70	73	71	59	67	-	-	-	72

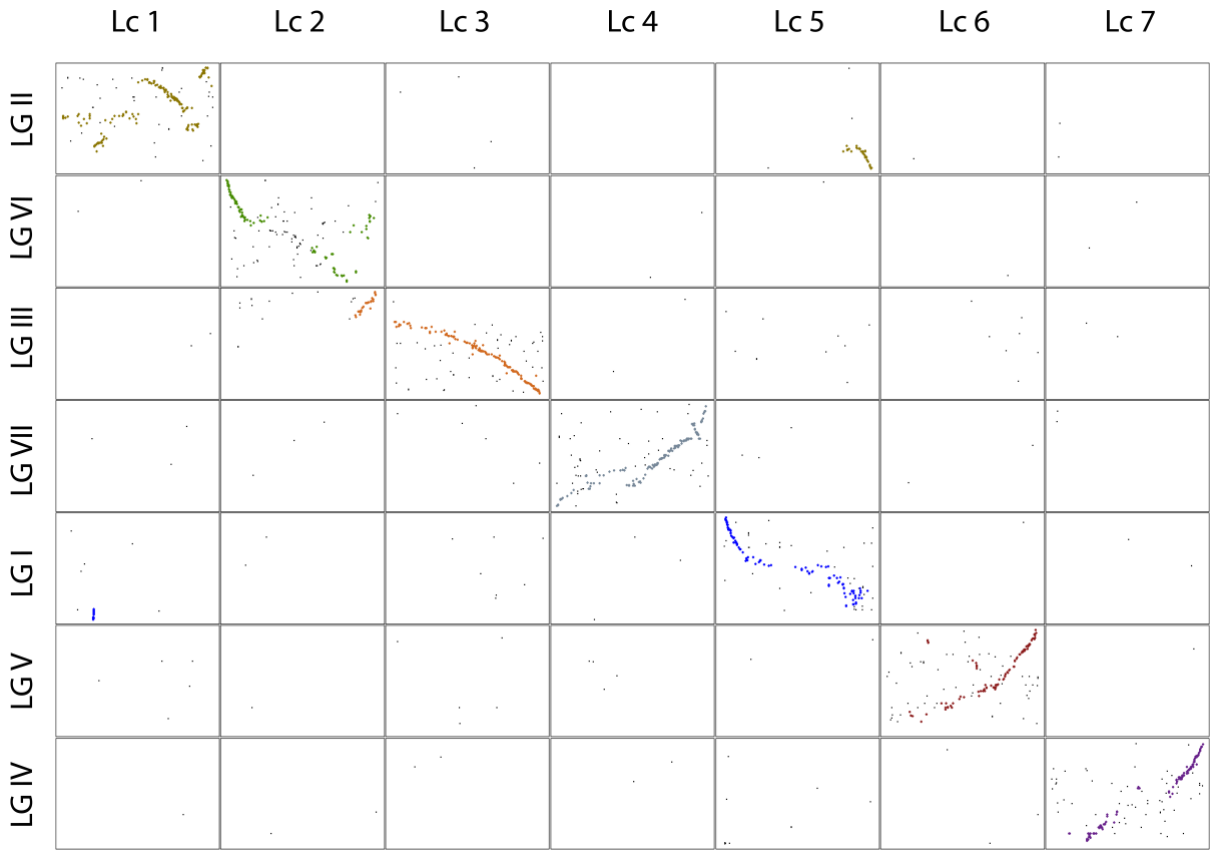


Figure 3. 4: Comparative analysis between pea and L. culinaris. Syntenic regions highlighted in colour

3.3.3.2.2 Medicago truncatula

M. truncatula belongs to the *Trifolieae* tribe, a sister tribe *Fabeae* which is estimated to have diverged around 16-23 MYA (Lavin et al., 2005). Consistent with previous reports (Schaefer et al., 2012, Tayeh

et al., 2015, Ma et al., 2017), our results showed PsLGI, II and V were collinear along its entire length with Mt5, 1 and 7 respectively, while the following LGs corresponded with more than one chromosome: PsLGIII with Mt3 and 2, PsLGIV with Mt4 and 8, PsLGVI with Mt2 and 6 and PsLGVII with Mt4 and 8 (see Figure 3.5). Varying levels of inversions and genomic rearrangements were reported in different LGs which appeared consistent with findings in Tayeh et al. (2015). PsLGVI showed greatest level of syntenic complexity, comprising a central region corresponding to Mt6 and two distal regions with similarity to sections of Mt2. A similar level of complexity was also found on PsLGI where an assortment of inversions and genomic rearrangements occurred throughout the length of the LG, which appeared consistent to *L. culinaris*. To a lesser extent a region on PsLGVII, corresponding Mt4 also showed also exhibited a complex syntenic relationship with numerous small rearrangements and inversion.

		Hits in *M. truncatula*											
		Mt1	Mt2	Mt3	Mt4	Mt5	Mt6	Mt7	Mt8	total	mapped	unmapped	Syntenic (%)
Pisum	LGI	31	24	23	27	276 (260)	13	21	22	524	437	87	59.5
	LGII	364 (331)	22	27	32	31	16	29	16	638	537	101	61.6
	LGIII	49	127 (95)	382 (332)	44	23	28	30	32	854	715	139	59.7
	LGIV	41	39	30	213 (186)	42 (12)	21	25	167 (144)	701	578	123	59.2
	LGV	22	17	31	32	16	18	290 (270)	24	537	450	87	60
	LGVI	28	202 (169)	37	46	21	103	38	24	606	499	107	45.9
	LGVII	37	31	22	284	35	12	31	145	709	597	112	62.3
	Total	572	462	552	678	444	211	464	430	4569	3813	756	-
	Syntenic (%)	58	57	60	63	63	28	58	62	-	-	-	58.3

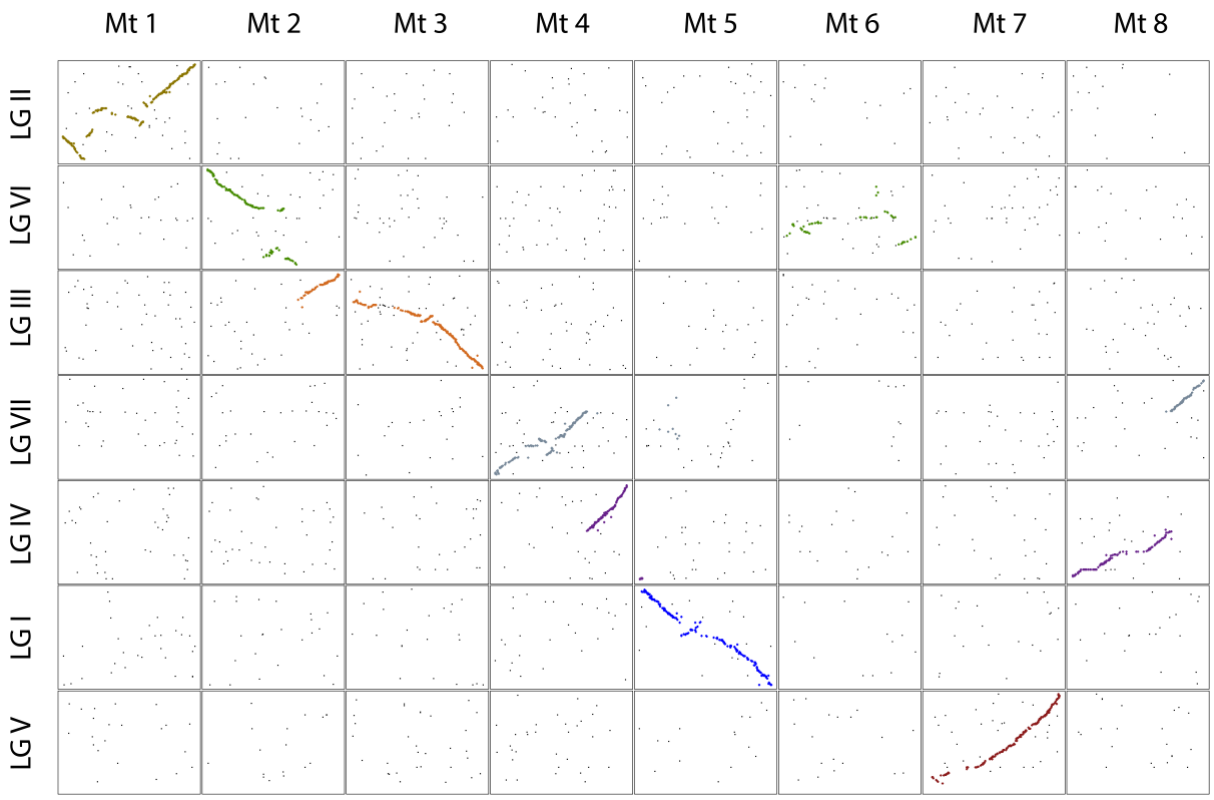


Figure 3. 5: Comparative analysis between pea and M. truncatula Syntenic regions highlighted in colour

3.3.3.2.3 Trifolium pratense (Red clover)

Like *M. truncatula*, *T. pratense* belongs to the *Trifolieae*, a sister tribe to the *Fabeae*. As shown in Figure 3.6, *T. pratense* demonstrated poor syntenic relationship in relation to its phylogenetic distance to pea. This was shown by its low proportion of mapped syntenic markers (42.53%) and a high-level of fragmentation (Table 3.7) and genomic rearrangements. This poor syntenic relationship is likely artificial, resulting from a poor establishment of the *T. pratense* genome. This is consistent to its high proportion of its genome (53.7%) being unassigned into pseudomolecules (detailed in Figure 3.1).

		Hits in *T. pratense*										
		Tr1	Tr2	Tr3	Tr4	Tr5	Tr6	Tr7	Total	mapped	unmapped	Syntenic %
Pisum	LGI	50	137 (64)	69	82 (33)	31 (11)	46	45	524	460	64	23.48
	LGII	251 (140)	41	56 (6)	49	24	57 (11)	35	638	513	125	30.6
	LGIII	53 (6)	90 (27)	159 (67)	76	33	52	250 (157)	854	713	141	36.04
	LGIV	75 (15)	130 (55)	62 (9)	97 (68)	100 (47)	56 (8)	59	701	579	122	34.89
	LGV	53	48	58 (6)	33	23 (6)	186 (129)	48	537	449	88	31.4
	LGVI	62	135 (37)	59	58	38	40	92 (11)	606	484	122	9.92
	LGVII	54 (8)	79 (29)	230 (169)	128 (56)	19	45	61	709	616	93	42.53
	Total	165	89	128	216	114	183	248	4569	3814	755	-
	Syntenic %	28.26	32.12	37.09	30.02	23.88	30.71	28.47	-	-	-	30.81

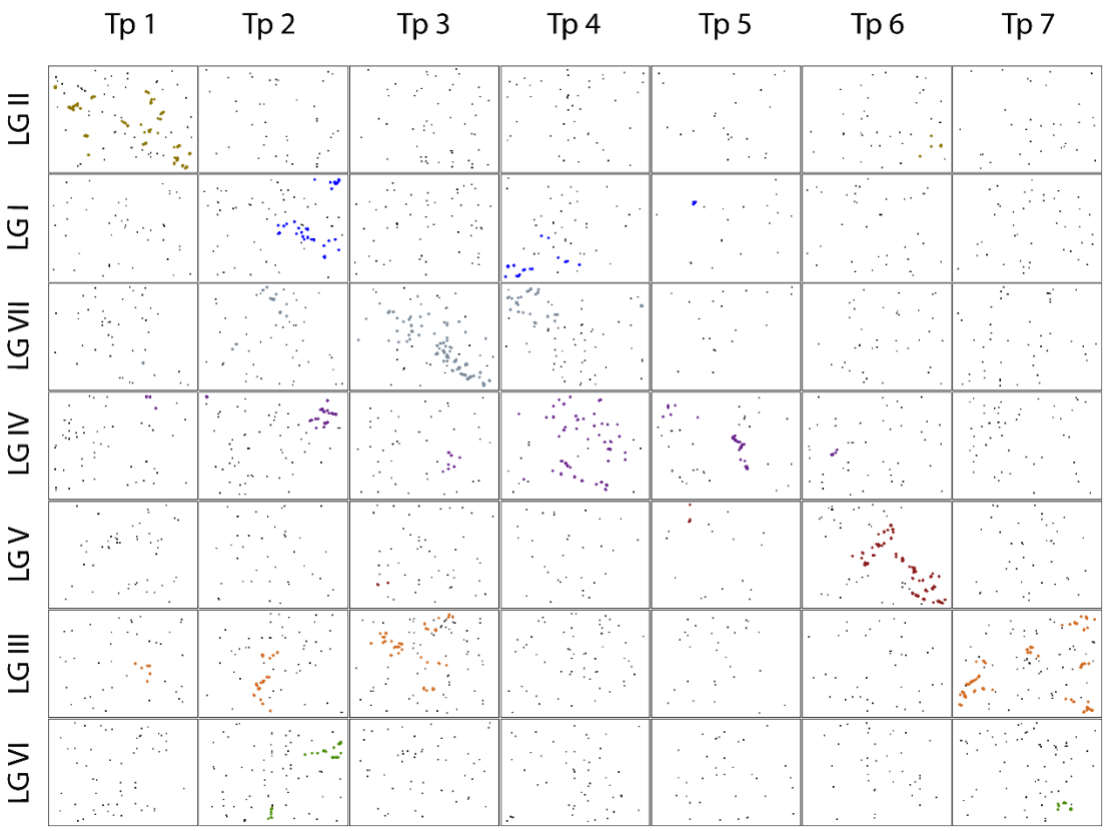


Figure 3. 6: Comparative analysis between pea and T. pratense. Syntenic regions highlighted in colour

3.3.3.2.4 Cicer arietinum (Chickpea)

Chickpea belongs to the tribe *Cicereae* which diverged from the *Fabeae* and *Trifolieae* tribes approximately 25-29 MYA (Choi et al., 2004). The syntenic relationship identified in this study was similar to that previously reported in Tayeh et al. (2015) with *C. arietinum* showing a slightly higher level of inversions and rearrangements than *M. truncatula* but overall showing close synteny (Figure 3.7).

PsLGII and VII were collinear along their entire length to *C. arietinum* chromosomes (Ca)4 and 6 respectively but included several major inversions. Interestingly, *M. truncatula* had a translocation event on PsLGVII which is absent in the corresponding *C. arietinum* region, indicating a genomic rearrangement specific to the divergence of *M. truncatula*. In all other LGs collinearity was shared between two or more chromosomes. PsLGI corresponded to Ca2 and 8, PsLGIII to Ca1 and 5 with possible novel regions on Ca4 and 6, PsLGIV to Ca7 and small novel region on Ca8, PsLGV to Ca3 and a small novel region on Ca2 and potentially on Ca4 and PsLGVI to Ca1 and Ca2. A minor translocation on PsLGIV with Ca1 was shown in Ma et al. (2017), but this was not identified in this study.

		Hits in *C. arietinum*											
		Ca1	Ca2	Ca3	Ca4	Ca5	CA6	Ca7	Ca8	total	mapped	unmapped	Syntenic (%)
Pisum	LGI	36	145 (129)	21	31	36	34	38	107 (94)	524	448	76	49.78
	LGII	28	27	34	348 (314)	26	45	29	16	638	553	85	56.78
	LGIII	116 (75)	32	27	61 (12)	343 (295)	77 (16)	46	20	854	722	132	55.12
	LGIV	47	28	33	44	42	38	328 (289)	33 (12)	701	593	108	50.76
	LGV	25	30 (10)	265 (235)	33 (7)	33	38	27	14	537	465	72	54.19
	LGVI	196 (153)	57 (25)	52	30	39	55	48	40	606	517	89	34.43
	LGVII	37	35	26	45	45	367 (333)	33	17	709	605	104	55.04
	Total	485	354	458	592	564	654	549	247	4569	3903	666	-
	Syntenic (%)	47	46	51	56	52	53	53	42.9	-	-		50.87

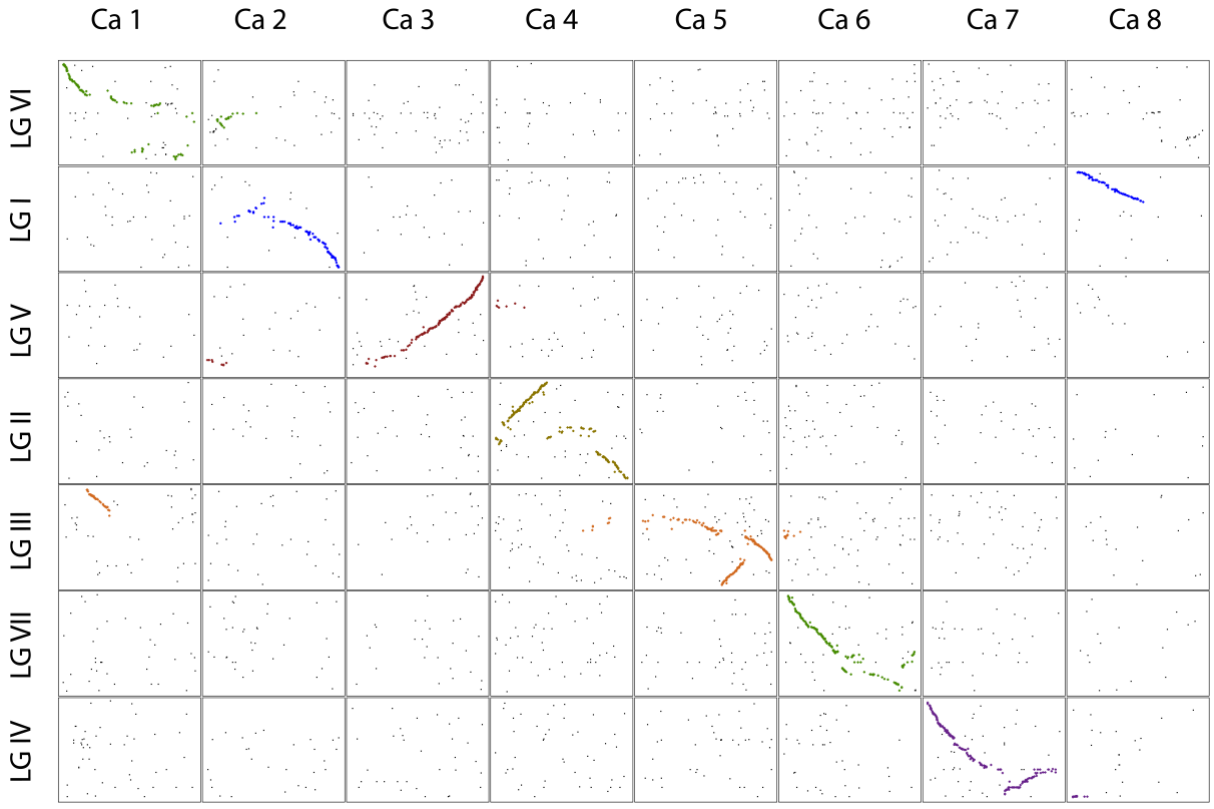


Figure 3. 7: Comparative analysis between pea and Cicer arietinum. Syntenic regions highlighted in colour

3.3.3.2.5 Phaseolus vulgaris (Common Bean)

P. vulgaris, belongs to the *Phaseoloid* clade and as such is the species most distantly related to pea that has been used in this comparative study. In general, and as expected the overall level of collinearity was much lower in this comparison (Figure 3.8). This consisted of many more fragmented syntenic blocks (Table 3.7) which is consistent with previous descriptions (Lee et al., 2017, Ma et al., 2017). Major features detected included a moderate level of synteny on Pv4 and 11 corresponding to PsLG1 and 5 respectively, as well as large syntenic blocks within PsLGIV and VII with Pv9 and 10 respectively.

		Hits in *P. vulgaris*														
		1	2	3	4	5	6	7	8	9	10	11	total	mapped	unmapped	Syntenic (%)
Pisum	LGI	17	13	25	88(65)	27	31	13	28	49	94(58)	30	524	415	109	29.6
	LGII	28	25	38	35	141(97)	27	26	24	50(7)	36	78(45)	638	508	130	29.3
	LGIII	47	28	110(69)	61	39	53(24)	83(59)	33	56	56	83(46)	854	649	205	30.5
	LGIV	35	64(53)	48	39(8)	28	36	26	24	155(131)	39	38	701	532	169	36.1
	LGV	28	21	26	74(58)	35	40(19)	20	27	28	26	96(70)	537	421	116	34.9
	LGVI	22	29	56(32)	43	48	62(47)	38(14)	62(28)	40	40(8)	37	606	477	129	27
	LGVII	78(44)	54(11)	28	54 (8)	35	23	32	29	59	138(92)	37	709	567	142	27.3
	Total	255	234	331	394	353	272	238	227	437	429	399	4569	3569	1000	-
	Syntenic (%)	17	27	31	35	27	33	31	12	31.6	36.8	40	-	-	-	30.7

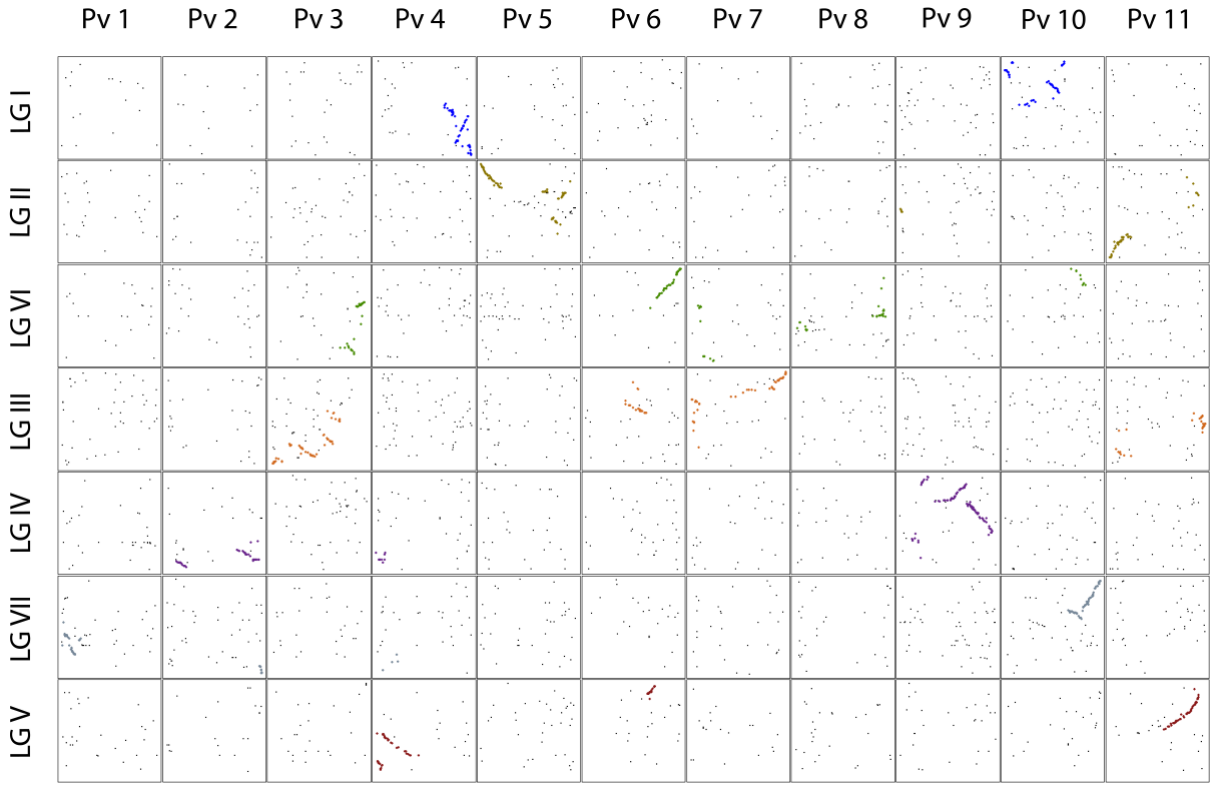


Figure 3. 8: Comparative analysis between pea and P. vulgaris. Syntenic regions highlighted in colour

3.4 Discussion

The generation of the 4.6K high-density, high-confidence linkage map reported in this chapter is an essential first step to investigating the genetic control of domestication in pea. Apart from the map of Tayeh et al. (2015), which was published during the course of this present study the map developed in this chapter is the next most densely populated linkage map for *P. sativum* currently available.

3.4.1 Map Quality

A high quality map is imperative for any molecular study (Darvasi et al., 1993) and depends in part on marker density, while accurate ordering and coverage of markers is important for candidate gene searches and high-resolution mapping (Collard et al., 2009, Liu, 2017). This study was provided with a large DArTseq genotyping dataset from a wild x domesticated pea population. This consisted of 137 individuals thereby exceeded the arbitrary number of 50 minimum individuals required for QTL analysis (Young, 1994). Due to computation requirements, conventional mapping techniques were unable to map such a large dataset. This is because the number of possible marker positions increases exponentially as the number of markers increases, and as a result, large marker datasets can increase the requirement for computational power and potentially introduce significant uncertainty in marker ordering (Collard et al., 2009, Liu, 2017).

To resolve this issue a *skeleton* map was initially developed using a marker binning process, reducing marker numbers from 6263 markers to 2091 with minimal effect to genome coverage. Remaining markers were later inserted back into the skeleton map to significantly increase mapping efficiency with minimal impact to marker order (Jighly et al., 2015). Genotyping errors can cause falsely separated, exaggerated distances and incorrect marker order (Hackett, 2002, Lehmensiek et al., 2005, Close et al., 2009), therefore removal of these lower quality markers has shown to improve both accuracy and map robustness. In this study all markers were assessed on their quality using a specified quality rank criterion. Lowest quality bin markers were removed during the development of the *skeleton* map as this significantly reduced map stability. To ensure map robustness, an iterative mapping technique with a stringent marker expulsion criterion was employed, resulting in a high-confidence *skeleton* map consisting of 905 markers.

After inserting the remaining markers back into the Skelton map, the finalised map consisted of 4599 markers, making this the second most densely populated pea linkage map available. The total map was calculated at 1617cM, averaging at around 3 markers per cM equating to 0.35cM between adjacent markers. Markers distribution was relatively uniform across the seven linkage groups with no adjacent markers exceeding 10cM. The maximum distance was reported as 5.80cM, which was within the proposed <20cM QTL mapping detectability threshold (Tanksley, 1993, Holland, 2007). Map

assembly was assessed by comparing relative size of linkage groups to previous studies. Results were relatively consistent with previous linkage maps (Weeden et al., 1998, Duarte et al., 2014, Sindhu et al., 2014, Sudheesh et al., 2015, Tayeh et al., 2015, Ma et al., 2017) and cytogenetic maps (Ellis and Poyser, 2002) therefore giving confidence to the assignment of markers per linkage group, although LGV was noticeably larger than the consensus. This may be caused by lower quality markers exaggerating its size which is supported by the comparatively lower marker density visualised in Figure 3.3. However, no segregation distortion was detected in any markers within this linkage group therefore indicating genotyping quality was good.

A high level of linkage distortion was reported in LGVII and moderate levels to LGII, III and IV, however because the distorted markers occurred in discrete regions rather than showing random distribution (data not shown), we considered that this was unlikely caused by low mapping quality. Linkage distortion is common phenomenon in many plant systems (Liu et al., 2010) but is particularly prevalent within biparental RIL population (Yamagishi et al., 2010), such as the one used in this study. This is because alleles can frequently be lost in wild populations (Liu et al., 2007) with greater concentration around domestication related regions (Blair et al., 2018). In pea, Tayeh et al. (2015) observed commonalities to linkage distortion in LGII and LGVI across multiple wild x cultivated populations. As to whether corresponding segregating distortion are common across different population is unknown (Liu et al., 2010) however segregation distortion reported here was not consistent with Tayeh et al. (2015).

3.4.2 Synteny with closely related species

To determine the quality of our map assembly, comparative analysis was conducted on the well-established *M. truncatula* and *C. arietinum* genomes and compared with previous detailed comparative analysis conducted by Tayeh et al. (2015). In addition, comparative analysis was conducted on outgroup *P. vulgaris* to allow assessment of synteny with a comparatively large phylogenetic distance. Currently there are no detailed comparative analysis of pea with *T. pratense* and *L. culinaris* and therefore these were analysed to determine suitability as a reference genome for functional genetic studies.

With the exception for *T. pratense*, all five species shared synteny with pea in proportion to phylogenetic distance. The poor syntenic relationship with *T. pratense* was thought to be artificial and likely caused by the poor establishment of its reference genome, therefore resulting in false positive syntenic fragmentation. This was supported by previous comparative analysis between *T. pratense* and *M. truncatula* which also showed low level of synteny (Jan et al., 2014, De Vega et al., 2015, Ištvánek et al., 2017), despite both species belonging to the *Trifolieae* tribe.

Comparative analysis with the well-established *M. truncatula* and *C. arietinum* genomes were consistent to previous reports therefore supporting the assembly of our map (Bordat et al., 2011, Sindhu et al., 2014, Tayeh et al., 2015, Ma et al., 2017). From our results it was possible to highlight the complex assortment of inversions and rearrangements in PsLGVI, which has constrained functional genetic studies in this region. This study also illustrated moderate level of inversions in PsLGII and to a lesser extent PsLGVII. The recent release of the ultra-high linkage map developed by Tayeh et al. (2015) also showed these regions exhibited complicated syntenic relationship, therefore providing further support to our map assembly. Our study identified three to five small novel translocations between *C. arietinum* on PsLGIII, IV and V which had not previously been reported.

As previously mentioned, this study has performed the first comparative analysis between pea and recently sequenced *T. pratense* and *L. culinaris* genomes. This was of significance due to their close phylogenetic affiliations, therefore potentially offering an alternative reference genome to *M. truncatula* and *C. arietinum* which are most commonly used in pea functional genetic studies. From our results, a large proportion of markers mapped to the *T. pratense* genome, however the proportion of syntenic markers was surprisingly low (30.8%) considering their relatedness. This was significantly lower than *M. truncatula* (58.3%) which belongs to the same tribe as *T. pratense* and significantly lower than *C. arietinum* (50.9%) which is more distantly related. This poor syntenic relationship is likely artificial and is believed to be caused by poor establishment of the *T. pratense* genome, which is supported by the large proportion (53.7%) of its sequence unassigned to its genome (Jan et al., 2014, Ištvánek et al., 2017). In light of these findings it is determined the current assembly of *T. pratense* genome is too fragmented and considered unreliable for functional genetic studies. For *L. culinaris*, our results showed a comparatively low proportion of markers mapping to this genome, however based on the proportion of syntenic markers (72%) and size of syntenic blocks, this genome showed a greater level of synteny than *M. truncatula* and *C. arietinum*. This implicates *L. culinaris* as being highly informative and potentially particularly useful in regions where mapping has previously been constrained by poor synteny like PsLGVI, as previously described. It should be noted that the resolution of markers in this model is lower than other comparative maps and therefore determining the microsyntenic configurations cannot be as confidently established. Moderate levels of inversions and rearrangements were reported in PsLGII as previously found in other genomes, thereby suggesting this occurred with the divergence of pea. In contrast, PsLGVII appears to show stronger collinearity than the more distant related *M. truncatula* and *C. arietinum*, therefore inferring that translocation occurred before the divergence of pea and lentil. From these results it can be determined that *L. culinaris* genome will make an important addition to the pea genomic resource

repertoire, however it is recommended to be used alongside the more established *M. truncatula* and *C. arietinum* genomes due to its reduced marker coverage.

4 Conclusions

In the absence of a fully sequenced pea genome, molecular studies have had to rely upon synteny with closely related species. The development of this high density, high confidence linkage map for the NGB5839 x JI1794 RIL population will provide a framework for use of the for QTL and other genetic analyses in the following chapters. Comparison with other pea maps indicate the map is essentially complete and has no major structural differences. Comparisons with the physical maps of several related legume species validated previous findings of reduced synteny with increased phylogenetic distance (Phan et al., 2006); but also illustrated how, in poor syntenic or low genomic coverage regions, more distantly related species can be more useful. This is the first study to compare genomic structure between the recently released *T. pratense* and *L. culinaris* sequenced genomes. From this it was determined that *L. culinaris* would constitute a valuable resource for future pea genetic studies, while the quality of *T. pratense* genome was insufficient for meaningful comparisons. Overall, the work presented in this chapter has made a significant addition to existing genomic resources for molecular research in pea and will be used to infer evolutionary process of pea domestication and identify potential candidate genes in later chapters.

Chapter 4: QTL analysis between a wild x domesticated *Pisum sativum L.* cross

4.1 Introduction

Common selective pressures for increased productivity and improved harvest efficiency have resulted in a convergent evolution of a common suite of agronomic traits, often referred to as the *Domestication Syndrome* (Hammer, 1984). These changes include free germination, loss or reduction of seed dispersal mechanisms and floral inductive requirements, increased self-fertilisation, alterations to plant architecture, enlargement of edible parts, reduction in toxicity and increased palatability (Doebley et al., 2006). It is generally agreed that pod indehiscence and loss of seed dormancy were the two critical traits for domestication, while changes in other traits probably occurred after domestication, during the diversification phase (Abbo et al., 2011, Abbo et al., 2014). Understanding the underlying genetic components of domestication-related traits has been of interest to both evolutionary geneticists and plant breeders, offering a unique window into its evolutionary history as well as providing a platform for further crop improvement efforts.

Pea was one of the earliest domesticated plants and today is a globally important food crop (FAOSTAT, 2018), yet our understanding of the underlying genetic changes incurred is limited (Weeden, 2007). The critical domestication trait of reduced pod dehiscence is generally considered to be mainly under the control of the *Dpo1* locus (Blixt, 1972, Hradilová et al., 2017). This was found to be true in both domesticated pea lineages *P. sativum var sativum* and *P. sativum var abyssinicum* (Weeden, 2007) indicating an early evolutionary adaptation, therefore supporting this as a critical domestication trait. Additional QTLs influencing pod dehiscence have also been identified (Weeden et al., 2002, Weeden, 2007) but unlike *Dpo1* these do not appear to have been universally adopted across the domesticated germplasm. Less is known about the loss of seed dormancy, also recognised as a critical domestication trait (Radchuk and Borisjuk, 2014, Smykal et al., 2014). Transcriptome analysis showed a down regulation of phenylpropanoid and flavonoid biosynthesis genes in non-dormant lines (Hradilová et al., 2017) indicating a possible link between pigmentation and dormancy. This was supported by Weeden (2007) who identified a QTL located near the Mendel *A* gene, which controls anthocyanin formation in flowers, stems and seeds. However, the existence of numerous pigmented lines within the domesticated germplasm suggests that this was not a critical modification and therefore not responsible for the primary dormancy loss associated with domestication (Smykal et al., 2014, Hradilová et al., 2017). For flowering time, five naturally variant loci have identified which include *Hr*, *Sn*, *LF* and *E* (Murfet, 1971, Murfet, 1973, Weller et al., 1997) and *QTL-V* (Weeden, 2007). *Hr* has been identified as *ELF3*, with mutant alleles causing a substantial reduction in photoperiod responsiveness.

Distribution of the mutant *hr* allele across the germplasm has suggested its possible role in the expansion of pea into Northern Europe (Weller et al., 2012). The *Sn* locus further reduces the photoperiod sensitivity and encodes the circadian clock-related protein LUX, with mutant alleles found within a discrete population carrying the *hr* allele, indicating a more recent origin (Liew et al., 2014). The *LF* locus has been identified as a *TERMINAL FLOWER 1* (*TFL1*) homolog (*TFL1c)* with numerous naturally occurring mutants reported (Murfet, 1985, Foucher et al., 2003). The *E* and the *QTL-V* loci have not been characterised and their distribution across the germplasm is unknown. A recent examination of flowering time in short-day conditions using the same cross as this study identified two QTLs (Weller et al., 2012); one identified as *Hr* and a second on LGVI which may correspond to the *E* locus.

With the advent of high throughput genotyping technologies, it has become more feasible to explore complex traits. This has been achieved by using either linkage mapping or genome-wide association studies (GWAS). Linkage mapping identifies genomic regions affecting traits of interest in a segregating progeny (usually from a biparental cross) by correlating trait value and genotype, while GWAS measures genetic variability across a natural population to identify disequilibrium between genetic and phenotypic variation. Both techniques have their advantages and disadvantages, with linkage mapping presenting a high detectability but limited scope, whereas GWAS offers a wider scope but frequently, lower detectability (Ross-Ibarra et al., 2007, Mackay et al., 2009, Tang et al., 2010). The current lack of a genome sequence for pea means that a GWAS approach is not yet possible so this study adopted for a linkage approach, using the RIL population for which a high-density linkage map was developed in Chapter 3. This approach has been applied in the dissection of domestication related traits in numerous crop species (Olsen and Wendel, 2013a) and was utilised in the only other significant examination of this question in pea (Weeden, 2007).

This study intends to confirm and possibly extend this previous analysis (Weeden, 2007) and aims to achieve the most comprehensive, high resolution genetic dissection of a wild x domesticated cross in pea to date. As pod dehiscence and the *Dpo1* locus are already under investigation elsewhere, this study has primarily focused on the genetic analysis of two key domestication traits; loss of seed dormancy, which has been ascribed as a critical domestication (Abbo et al., 2014) and flowering time, a diversification trait that has enabling pea to grow outside its normal geographical range (Weller et al., 2012). This study will also analyse the genetic control of additional domestication syndrome traits and other morphological changes to gain a broad picture of changes to the genetic architecture that have accompanied domestication and improvement. Findings from this study should provide a platform for future fine mapping studies and identification of specific causal genes.

4.2 Methods

4.2.1 Plant material and growth conditions

4.2.1.1 Plant material for QTL analysis

An F8+ RIL population consisting of 137 individuals was developed by single seed descent from the F2 population of the biparental cross between a wild (JI1794) and modern cultivar (NGB5839), described in Weller et al. (2012), (J Vander Schoor and J Weller, unpublished).

For each line four seeds were sown except for SH37, SH48, SH90, SH95, SH104 and SH130 which were known for their poor survival rate and therefore 6 to 8 seed were grown. All plants were grown under controlled glasshouse long-day (LD) conditions between December 2014 and February 2015.

4.2.1.2 Additional populations

To validate specific QTLs advanced generation segregating populations or Near Isogenic Lines (NILs) were used. Advanced generation segregating population were developed by preferentially selecting and advancing lines from the parental F2 population, making use of 81 gene-based markers that had been genotyped in this population (Weller et al. 2012, V. Hecht et al., unpublished; see chapter 2 for details). The NIL populations had previously been developed where the wild dominant *HR* allele from the WL1771 line was introgressed into the domesticated NGB5839 background through six successive backcrosses (Weller et al. 2012).

4.2.2 Phenotyping

A total of 49 individual trait measurements were made across several categories: pigmentation, dormancy, flowering time, plant architecture, productivity, development rate and leaf traits. (Tables 4.1 and 4.2).

Table 4. 1: Details of traits measured across RIL population.

Trait category	Measurement	QTL no.	QTL	Description
Pigmentation	Total flower colour	1	Fc	The presence or absence of floral pigmentation
	Green testa	2	Gt	Presence or absence of green testa (1 = present, 0 = absent)
	Red testa	3	Rt	Presence or absence of red testa (1 = present, 0 = absent)
	Mottled	4	Mt	Subjective rank score between 0 and 2 of testa mottling where 0 exhibits no mottling and 2 is highly mottled.
	Black spots	5	Bs	Presence or absence of black spots (1 = present, 0 = absent)
	Black hilum	6	BH	Presence or absence of black hilum (1 = present, 0 = absent)
Dormancy	Permeability	7	Perm	Time duration for desiccated seed to fully imbibe. This was measured using 5 representable seeds fully submerged in petri dish of water kept at a constant 22oC. This was conducted 1 to 6 months after harvesting of parental line to ensure seeds were fully desiccated but also maintain seed quality. Seeds were classified as fully imbibed after seed weight gain plateaued after initial uptake of water. Seeds were measured at timed intervals using A&D weighing GR-200 lab balance. For each measurement, seeds were removed from petri-dish and surface water around testa removed using A&D weighing GR-200 lab balance. Log transformation was applied so that values data fitted a normal distribution.
	Testa thickness Digital Calipers	8	TTd	Circular section of the testa is removed using a razor blade, avoiding the hilum. The circular section is then cut into equal quarters with thickness measured twice on three of the four sections using a digital caliper and micrometer respectively. Where possible 5 independent seeds were measured, totalling 30 measurements per RIL line (2*3*5 = 30).
	Testa thickness - Micrometer	9	TTm	
	Testa roughness	10	GTY	Subjective ranking of testa roughness from 1 (smooth seeded) to 5 (high testa roughness).
	Seed desiccation	11	Sd	Relative increase in weight from desiccated seed to fully imbibed. Seed weights were collected a minimum of 30 days after harvest and with protrusion of radicle.
FT	Flowering time	12	DFT	Number of days between sowing and the opening of the first developed flower
	Floral node	13	FN	Number of nodes on the main stem to first flower is developed
Plant architecture	Plant height at 22 days	14	ph	Length of stem from first node to apex. Measurement were taken from top of pot using a ruler, 22 days from sowing (mm)
	Total plant Height	15	PH	Length of stem from first node to apex at end of life cycle, measurement collected by removing plant from pot (mm)
	Relative branch length at 22 days	16	br	Cumulative branch length from nodes 1 and 2. Branch length is measured 22 days after sowing and divided by plant height at 22 days
	Total relative branch length nodes 1 and 2	17	BR	Cumulative branch length from nodes 1 and 2 divided by plant height at end of life cycle
	Internode length between 3 and 6	18	INL	Internode length between 3rd and 6th node on main stem Measurement collected at end of life cycle
	Internode length between 6 and 9	19	INM	Internode length between 6th and 9th node on main stem. Measurement collected at end of life cycle
	Internode length between top 3 nodes of plant	20	INT	Internode length between apex and 3rd node down. Secondary growth not included in measurement. Measurement collected at end of life cycle
	Lower stem thickness	21	Lst	Average width of stem between 3rd and 4th node from base of plant. Plants scored using digital caliper with three independent measurements taken per plant. Measurements collected prior to plant senescing
	Upper stem thickness	22	Ust	Average width of stem between 3rd and 4th node from top of plant, excluding secondary growth. Plants scored using digital caliper with three independent measurements taken per plant. Measurements collected prior to plant senescing
	Length of petiole at node 7	23	PLi	Length of petiole from mainstem to first set of leaflets. Measurements collected at end of plant life cycle.

	Length of petiole at node 13	24	PLii	Length of petiole from mainstem to second set of leaflets (or tendrils if secondary set of leaflets have not developed). Measurements collected at end of plant life cycle.
	Primary pod axis	25	PPA	Refer to Figure 4.1. Measurements collected from first flowering node using a ruler at end of life cycle
	Secondary pod axis	26	SPA	
	Dual pod axis	27	DPA	
	Probability of dual pods	28	DPP	Proportion of plants with dual pods, where 1 = all replicates have dual pods, 0 = no replicates have dual pods

Productivity	Seed weight	29	Sw	Average dried seed weight of 10 representable seeds. Weight was scored after a minimum of 30 days from harvesting to ensure seed is fully dried
	Seed number	30	Sn	Total number of viable seeds produced
	Duration of plant growth	31	Gd	Number of days from sowing to apex terminating
	Reproductive nodes	32	Rn	Total number of floral nodes on mainstem
	Neoplasticity	33	Np	Subjective ranking of neoplasticity on node 10 where 0 = no neoplasticity to 3 = significant neoplasticity

Development rate	Node to change to 4 leaflets	34	Nc	Node at which four leaflets are expressed
	Number of nodes expanded at 13 days	35	Nei	Total number of nodes expanded after 13 days
	Number of nodes expanded at 41 days	36	Neii	Total number of nodes expanded after 41 days
	Number of nodes expanded at 56 days	37	Neiii	Total number of nodes expanded after 56 days
	Development rate between 7-13 days	38	DRi	Total number of nodes expanded between days 7 and 13 days
	Development rate between 36 and 41 days	39	DRii	Total number of nodes expanded between days 36 and 41 days
	Development rate between 49 and 56 days	40	DRiii	Total number of nodes expanded between days 49 and 56 days

Leaf traits	Leaf length	41	LfH	Height from base to the tip of the leaf collected at node 10. Leaf length was scored by ImageJ using scanned images
	Leaf width	42	LfW	Width of widest region on leaflet collected at node 10. Leaf width was scored by ImageJ using scanned images
	Relative height of leaf at widest point	43	HwT	Height at base to widest point of leaflet divided by total leaflet height. Leaflet collected from node 10, scanned then measured using software ImageJ
	Leaf roundness	44	FvT	Height divided by width of leaflet collected at node 10
	Leaf perimeter	45	LfP	Perimeter of leaflet at node 10 measured using ImageJ software from scanned image
	Leaf area	46	LfA	Area of leaflet at node 10 measured using ImageJ software from scanned image
	Leaf shape principle component 1	47	Pci	Leaf shape analysed using the program, *LAMINA (Bylesjö et al., 2008)*. Two principle component analysis were used, Principle component 1 appears to measure leaf width and Principle component two relative position of widest point of leaf, see Figure 4.2.
	Leaf shape principle component 2	48	Pcii	
	Leaf serrations	49	LfS	Subjective ranking of leaflet serrations between 0 and 2 observed from node 10, with 0 = no serrations and 2 = highly serrated

Table 4. 2: Details of any additional traits measured on additional populations not described in Table 4.1.

Trait	Description
Internode length 3 to 9	Distance between node three and node 9, measurements taken using ruler at end of life cycle.
Total number of nodes	Total number of nodes counted on main stem, evidence of secondary growth not included
Relative internode length	Number of nodes identified on main stem divided by total plant height
Seeds on main stem	Cumulative number of seeds collected from main stem only
Seeds from branches	Cumulative number of seeds collected from main stem only
Pods on main stem	Cumulative number of pods collected from main stem only
Pods from branches	Cumulative number of pods collected from main stem only
Node when Sum of Leaf and tendril ≥ 4	Node when the sum of leaflets and tendril is greater than or equal to 4, 6, 8, 11, 13 respectively
Node when Sum of Leaf and tendril ≥ 6	
Node when Sum of Leaf and tendril ≥ 8	
Node when Sum of Leaf and tendril ≥ 11	
Node when Sum of Leaf and tendril ≥ 13	
Development rate week 5	Difference in total number of nodes counted between weeks 4 and 5
Development rate week 9	Difference in total number of nodes counted between weeks 8 and 9

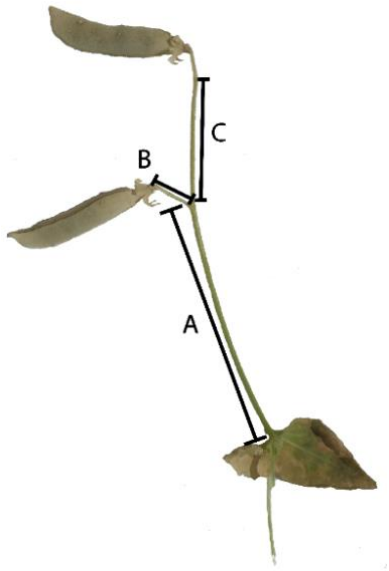


Figure 4. 1: Photograph illustrating measurements taken for pod architecture. A) Primary pod axis (PPA); distance between stipule and base of secondary pod axis B) Secondary pod axis (SPA); distance between base of primary pod axis to base of first pea pod and; C) Dual pod axis (DPA); distance between base of secondary pod axis and base of second pea pod.

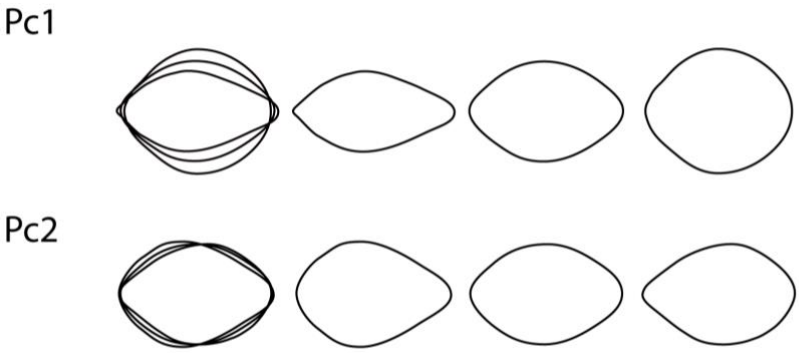


Figure 4. 2: Illustration of two Leaf shape characteristics identified using leaf shape measuring program LAMINA, Principle component 1 (Pc1) and principle component 2 (Pc2).

4.2.3 Statistical analysis

4.2.3.1 Phenotypic scoring

All replicated scores were averaged, and standard error calculated. Any plants exhibiting abnormal growth in relation to replicates of the same line were excluded.

4.2.3.2 QTL analysis

QTL analysis was performed using MapQTL 6 software (Van Ooijen, 2009) in conjunction with the high density map previously developed in Chapter 3. For computational efficiency the map was reduced from 4599 to 3073 markers by excluding every third marker. QTL searches were conducted by the following protocol. An initial search for QTLs was conducted using the *interval mapping* function, a method that estimates effect and position of QTL within two flanking markers (Lander and Botstein, 1994). QTLs were defined by a logarithm of odds (LOD) score with a significance threshold of 3, a standard statistical method used for calculating the probability of linkage between two loci. To determine the most significant marker, for each putative QTL identified, the marker with the highest LOD score together with four adjacent markers either side of peak marker were used in *Automatic Cofactor Selection* (*ACS*). Multiple successive searches for QTLs were performed using the *Multiple QTL Model* (MQM) function. This method increases the power of QTL analysis by reducing residual variances attributed to previously identified QTLs (cofactors). The *MQM* and *ACS* functions were reiteratively employed until all QTLs exceeding the specified threshold were identified. With each QTL search, cofactors are reviewed and adjusted if necessary to ensure they remain the most representative of the QTL.

The amount of variation explained by each QTLs was estimated using the *coefficient of determination* (R^2) which is represented as the *Phenotypic Variance Explained* (PEV). QTLs with a PEV score greater than 15% were considered as major and those less than 15% as minor. As the map had no gaps >10cM, QTLs that mapped $\leq$10cM apart were considered to be effectively indistinguishable.

QTLs were labelled according to trait descriptor and linkage group (e.g. the testa thickness QTL on LG2 as *TT2*). Distinct QTLs for the same trait on the same chromosome were distinguished by a suffix (e.g. flowering time QTLs on LG3 as *FT3a* and *FT3b*).

4.2.4 Figures

QTL figures were developed using the statistical software R-studio. LOD scores and relative linkage position were extracted from MapQTL by selecting the *Results charts* function. Linkage groups that did not have a significant QTL were not shown.

4.2.5 Identifying significant difference between genotypes

To determine whether a specific locus have significant on trait a Tukey's pairwise analysis was performed on advanced generation segregating populations ($P \leq 0.05$). This was conducted between Domesticated-Heterozygous (B-H), Wild-Domesticated (A-B) and Wild-Heterozygous (A-H).

4.3 Results

4.3.1 Morphological variations

Several of the distinct morphological differences between the RIL parent lines JI1794 and NGB5839 are illustrated in Figure 4.3. NGB5839 is a near isogenic line to cv *Torsdag*, carrying induced *le-3* mutant allele (Sponsel and Reid, 1992, Lester et al., 1999), and is a small erect plant with white flowers, non-dehiscent pods and large, non-pigmented non dormant seeds with thin, permeable testa s. Line JI1794 (*P. sativum var humile*) is a representative of the wild "northern humile" and is characterised by its tall twining growth habit, pronounced basal branching, purple flowers, dehiscent pods, and small, pigmented seeds with thick, impermeable testa.

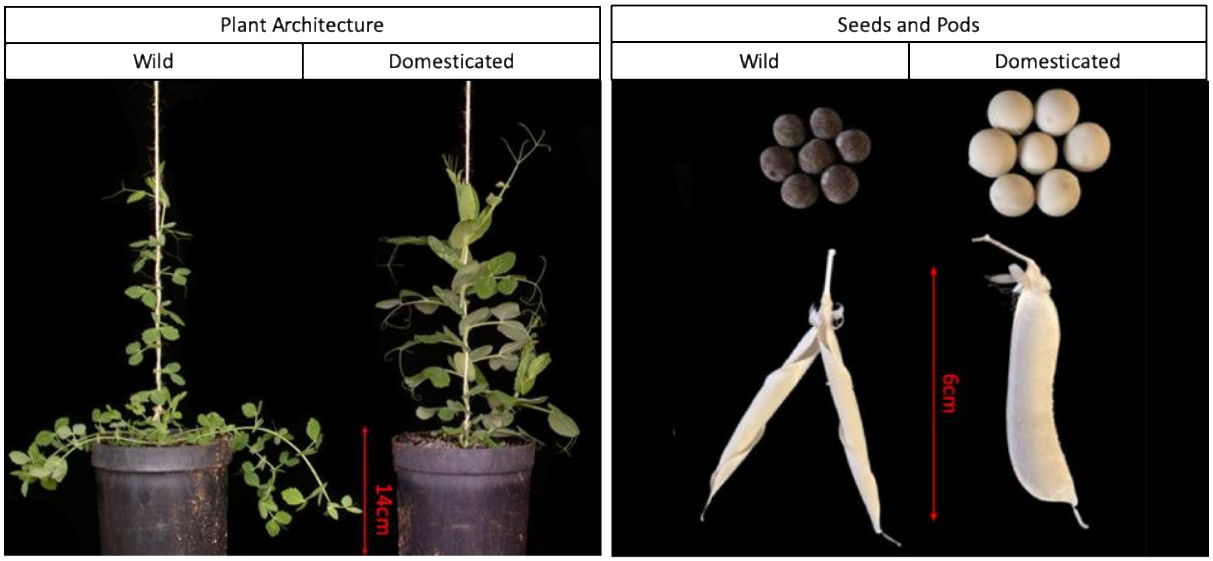


Figure 4. 3: Image of representative wild and domesticated lines showing phenotypic disparities including the plant architecture, seeds and pods

Based on prior knowledge of these parental lines, certain loci were known to be segregating within this population, including *Le, ELF3/Hr, QTL6/E, Dpo1, GTY, A and Pl* (Table 4.3). The classical Mendelian loci mottled testa (*M*) and purple testa flecking (*Fs*) would also be expected to segregate, although their expression would be masked to some extent in the anthocyanin-deficient *a* background. Causative mutations for *HR* (Weller et al., 2012), *A* (Hellens et al., 2010) and *Le* (Lester et al., 1999) have been characterised with actual positions mapped, and remaining loci have been mapped with varying levels of resolution but the causative genes are not yet known.

Table 4. 3: Details of previously identified Mendelian loci segregating within this population.

Loci	Prominent Trait	Position (LG)	Causative gene	Measured (Y/N)	reference
Dpo1	Dehiscent pod	III	-	N	(Hradilová et al., 2017)
GTY	Testa roughness	VI	-	Y	(Ellis et al., 1993, Weeden et al., 1998)
Hr	flowering time	III	ELF3	Y	(Weller et al., 2012)
QTL6	flowering time	VI	-	Y	(Weller et al., 2012)
Le	Plant height	III	Gibberellin 3P-Hydroxylase (*GA3ox1*)	Y	(Weeden et al., 1998, Lester et al., 1999)
Np	Neoplasticity	III	-	Y	(Weeden et al., 1998, Weeden, 2007)
A	Pigmentation	II	bHLH	Y	(Hellens et al., 2010)
M	Pigmentation	III	-	Y	(Weeden et al., 1998, Bordat et al., 2011)
Pl	Pigmentation	VI	-	Y	(Weeden et al., 1998, Bordat et al., 2011)
Fs	Pigmentation	V	-	Y	(Weeden et al., 1998, Weeden, 2006, Bordat et al., 2011)

A total of 49 different measurements (see Table 4.1) were scored and analysed across 135 of the original 137 individuals. Lines SH70 and SH77 excluded due to seed loss. A summary of measurements collected including phenotypic means and standard deviations trait are presented in Supplementary Figure 4.1.

4.3.2 QTL analysis

Across all 49 measurements, a total of 160 QTLs exceeding the LOD 3 significance threshold were identified with map positions illustrated in Figure 4.4 and detailed in Supplementary Figure 4.2. Between one and seven QTLs were detected for any given trait, with an average of just over 3 per trait. On average there were 22 QTLs per linkage group, but this distribution was not even across groups, with 4, 18, 55, 9, 15, 42, and 14 QTLs mapping to LGI-VII respectively. It is recognised that among the 49 traits measured several are likely to be highly correlated, with similar genetic control, therefore the 160 QTLs may represent a much smaller number of distinct loci. To take account of this the 49 measurements were grouped into the seven trait groups which include 1) pigmentation 2) seed dormancy related, 3) flowering time 4) plant architecture 5) productivity 6) development rate and 7) leaf morphology, (see Table 4.1). Where QTL of the same group was located within 10cM interval these were considered the same QTL. As shown in Table 4.4 and Supplementary Figure 4.3, total number of QTLs were recalculated at 89, however this again is likely to be an over representation as pleiotropic effects spanning more than one trait group were not accounted for.

Table 4. 4: Summary table of all QTLs involved per grouped domestication related trait. QTLs within a group which mapped within 10cM were considered the same QTL

Trait no.	Domestication trait	QTLs		
		A	B	Total*
1	Pigmentation	5	1	6
2	Seed dormancy	10	0	10

3	Flowering time	3	2	5
3	Plant architecture	3	13	16
5	Productivity	7	7	14
6	Development rate	7	5	13
7	Leaf morphology	9	16	25
Total				89

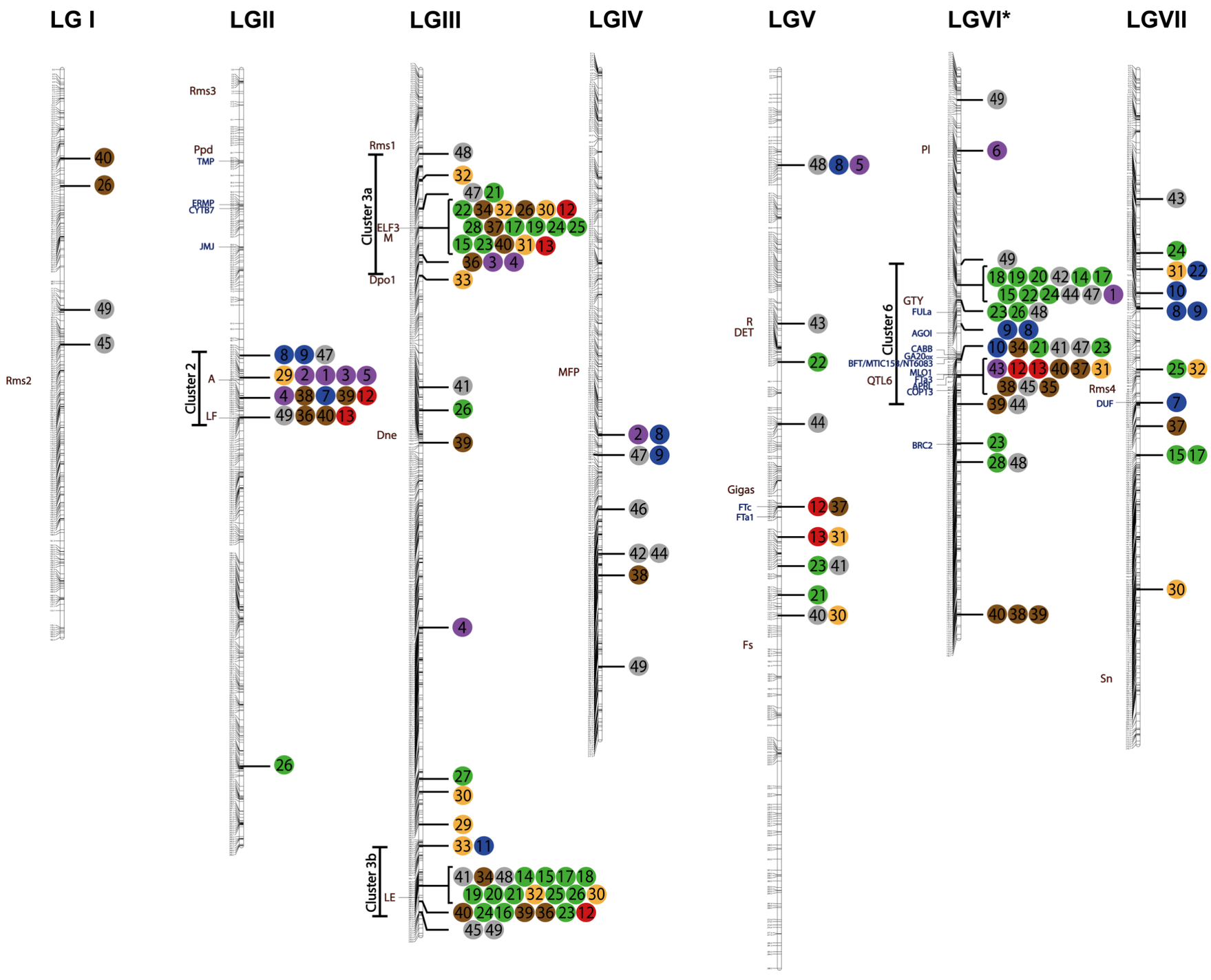


*Figure 4. 4: Linkage map, depicting mapped QTL position from measured domestication related traits. Relative markers positions are mapped on the left of each representative linkage group, brown text show relative positions of loci used in Bordat et al. (2011) except A, LF and Le which shows actual position. Blue text show anchor marks used in this study. QTLs are displayed to the right of each linkage group with numbers referring to trait number and colour to trait type. Trait types are: Purple – Pigmentation, Dark blue – Seed dormancy related, Red – Flowering time, Green – Plant architecture, Orange – productivity, Brown – Development rate, Grey – leaf traits * LGVI inverted to previous map. (high resolution version can be accessed in Cloudstor)*

4.3.2.1 Pigmentation (*A*, *M*, *Fs*, *Pl, Gt, Rt*)

In pea, most instances of pigmentation loss can be attributed to the loss of function of Mendel *A* gene, although other naturally occurring mutants are known to exist (Hellens et al., 2010). Presence and absence of pigmentation in flowers (Fc) was analysed to confirm mapping of *A*. Additional pigmentation traits were also analysed to confirm their segregation within population and potentially refine map positions relative to previous estimates This includes testa marbling (*M*), black/purple flecking (*Fs*), black hilum (*Pl*), red testa (*Rt*) and Green testa (*Gt*) (Figure 4.5).

As expected the loss of pigmentation *fc* mapped to Mendel *A* gene which will subsequently be referred to as *A*. All other measured pigmentation traits were epistatic to *A* except for black hilum also showed a QTL at the same position, consistent with their dependence on *A* function. The independence of *Pl* and *A* was consistent with findings in soybean where hilum colour could overcome the epistatic effect of pigmentation by the recessive *k1* locus (Yang et al., 2010). This was encoding as the *Argonaute5* (*AGO5*) mapped to Glyma.11G190900 (Cho et al., 2017), but despite this genetic similarity, this gene is not syntenic to the pea *Pl* locus in accordance to the Tayeh et al. (2015) linkage map.

Major QTLs for flecking, black hilum and marbling mapped to positions expected for *Fs*, *Pl* and *M* respectively (Weeden et al., 1998, Bordat et al., 2011), and an additional minor locus affecting marbling (*M3b*) was detected on LGIII. The red and green testa traits each demonstrated a Mendelian segregation in the presence of the functional *A* allele. The red testa trait co-segregated with the *M* locus, whereas the green testa trait (*GT4*) mapped to a novel region on LGIV distinct from previous reports (McCallum et al., 1997).

Trait code	Trait measured	QTL	LG	Position	Locus	Closest anchor marker	LOD	PEV	Associated genotype
1	Total flower colour	Fc2 (*A*)	II	99.627	A_1	A (0.00cM)	92.45	95.7	A
2	Green testa	GT2	II	98.906	3553850_3	A (-0.721cM)	4.15	6.8	A
2	Green testa	GT4	IV	119.542	5252051_1		19.27	41.7	A
3	red testa	RT2	II	99.627	A_1	A (0.00cM)	15.19	27.8	A
3	red testa	RT3	III	60.309	3548887_2	HR (-8.721cM)	14.03	25.1	A
4	Mottled	Mt2	II	105.354	3544617_2	Between A (+5.727cM) and LF (-6.646cM)	10.68	10.0	A
4	Mottled	Mt3a	III	60.309	3548887_2	HR (-8.721cM)	36.51	60.4	A
4	Mottled	Mt3b	III	179.179	5252081_3		3.02	2.4	B
5	Black spots	Bs2	II	99.627	A_1	A (0.00cM)	15.95	29.4	A
5	Black spots	Bs5	V	31.54	3550203_1		15.41	28.1	A
6	Black Hilum	BH6	VI	26.1	3568221_1		48.41	81.1	A

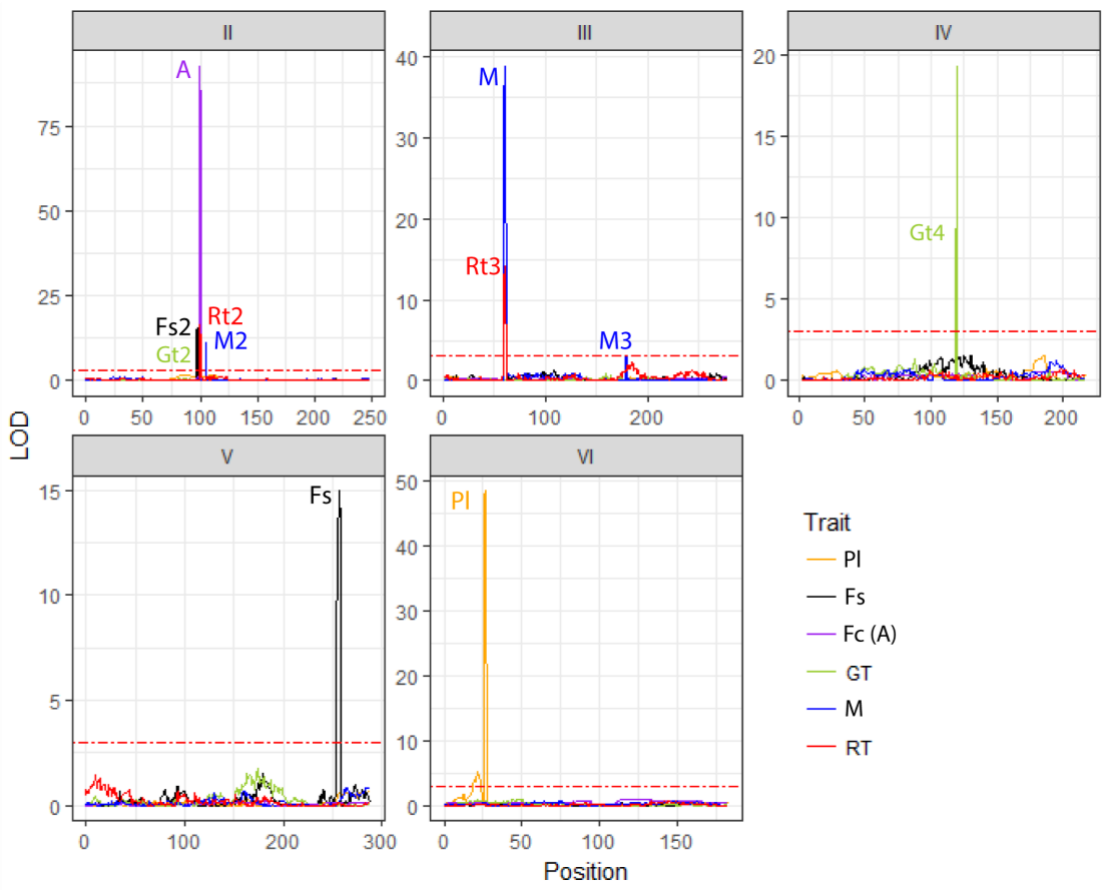


Figure 4. 5: QTL analysis results for pigmentation related traits a) Table detailing position, effect and allelic direction of all QTLs identified b) visual representation of all QTLs identified; Fc = presence/absence, M = Marbling, Fs = Blackspots, Pl = Black hilum, Gt = Green testa, RT = red testa. Presence of pigmentation mapped to Mendel *A* gene.

4.3.2.2 Seed dormancy

Loss of seed dormancy is recognised as a critical domestication trait that enables uniform germination rates and profitable harvesting returns (Abbo et al., 2011, Abbo et al., 2014). Hardseededness plays a prominent role in regulating dormancy in peas, but is poorly understood at the molecular level (Smykal et al., 2014). The permeability (*Perm*) trait was quantified as a direct measure of this dormancy, while testa thickness (TT), testa roughness (Sr) and seed desiccation (Sd) were also analysed to examine their potential contribution.

Testa thickness has been strongly implicated in seed dormancy and certain studies have even used it as a proxy for dormancy (Murphy and Fuller, 2017, Weeden, 2018). In this study testa thickness was measured using two different techniques, using either a digital caliper (*TTd*) or a micrometer (*TTm*) for greater robustness of the result. Testa roughness (*Sr*) was also considered a potential factor regulating seed dormancy as smooth seeded lines were ubiquitous across the cultivated germplasm (Zohary and Hopf, 1973, Zohary, 1989, Zohary et al., 2012), implying a tight association with

domestication. Lastly the testa desiccation (*Sd*) trait was considered as an indirect measurement of porosity.

As shown in Figure 4.6, permeability was controlled by two QTLs, including a major QTL on LGII (*Perm2*) and a minor on LGVII (*Perm7*). *Perm2* mapped near Mendel *A*, a result consistent with Weeden (2007) when analysed in a population segregating for *A*, but interestingly was absent in a primitive x cultivated population which incidentally was fixed for *A*. *Perm7* mapped to a region on LGVII distinct from the remaining four testa thickness QTLs as well as testa roughness and seed desiccation QTLs. These findings contrast to those of Weeden (2018) which provided evidence that seed dormancy was regulated by two testa thickness QTLs, which were both apparently different from *Perm2* and *Perm7*. Furthermore, Weeden (2007) also reported two other additional seed dormancy QTLs, one probably corresponding to Mendel *R* locus (round/wrinkled seed) on LG V and the second to the sticky seed (*S* locus) on LG II. However, neither of these traits were present in the NGB5839 parent line and therefore not relevant to our situation.

For testa thickness five QTLs were identified, this included major QTL on LGII (*TT2*) which as previously mentioned mapped near *Perm2* and *A*, as well as four novel minor QTLs on LGIV (*TT4*), LGV (*TT5*), LGVI (*TT6*) locus and LGVII (*TT7*). For testa roughness two QTLs were identified, a major QTL on LGVI corresponding to the *GTY* locus and a novel QTL on LGVII (Sr7). Interestingly both seed coat roughness QTLs mapped close to the testa thickness QTLs (i.e. *TT6* and *TT7*), raising the possibility that the genetic aspect for *TT6* and *TT7* may be the same as those regulating seed coat roughness *GTY* and *Sr7* respectively.

Trait code	Trait measured	QTL	LG	Position	Locus	Closest anchor marker	LOD	PEV	Associated genotype
7	Permeability	Perm2	II	107.1	5251908_1	Between A (7.473cM) and LF (-4.9cM)	11.11	32.7	A
7	Permeability	Perm7	VII	107.014	3542137_3	DUF (0.00cM)	3.43	8.6	A
8	TT digital caliper	TT2	II	91.765	3558561_1	A (-7.862cM)	9.76	16.6	A
8	TT digital caliper	TT4	IV	119.835	4663134_3	-	4.34	6.7	A
8	TT digital caliper	TT5	V	30.829	3560771_3	-	4.09	6.3	A
8	TT digital caliper	TT6	VI	82.22	3565980_1	Between FULa (+4.138cM) and AGOI (-2.369cM)	7.49	12.3	A
8	TT digital caliper	TT7	VII	76.167	3542996_3		7.1	11.5	A
9	TT Micrometer	TT2	II	91.765	3558561_1	A (-7.862cM)	14.55	29.1	A
9	TT Micrometer	TT4	IV	125.904	4655522_1	-	5.55	9.4	A
9	TT Micrometer	TT6	VI	81.644	3561639_2	Between FULa (+3.562cM) and AGOI (-2.945cM)	3.34	5.5	A
9	TT Micrometer	TT7	VII	76.707	4662724_3		3.5	5.7	A
10	Testa roughness	GTY	VI	91.222	4663141_3	Between AGOI (+6.633cM) and CABB (-0.468cM)	32.41	66.6	A
10	Testa roughness	Sr7	VII	71.888	3563615_3		3.9	4.6	A
11	Seed desiccation	Sd3	III	250.09	4656495_1	LE (+14.241cM)	5.85	21.1	A

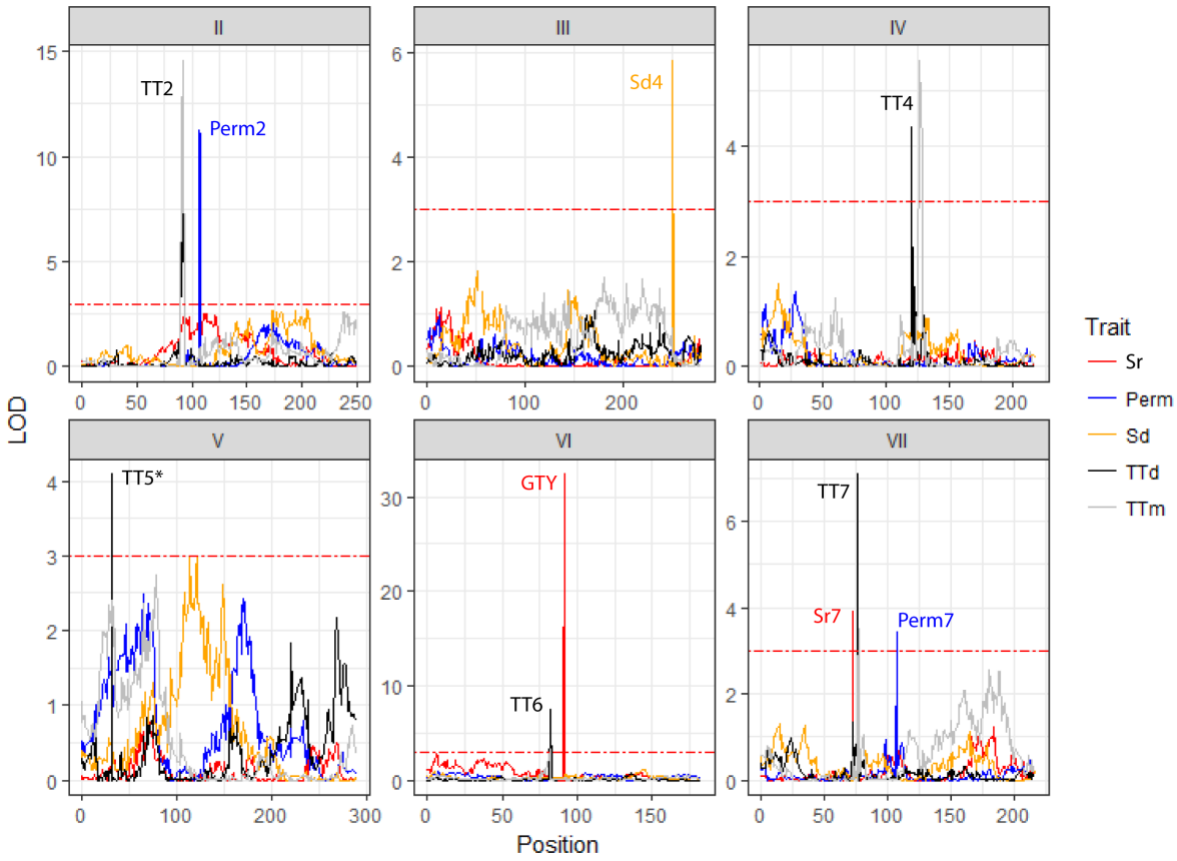


Figure 4. 6: QTL analysis results for Seed dormancy related traits a) Table detailing position, effect and allelic direction of all QTLs identified b) visual representation of all QTLs identified; Perm = Permeability, TTd/m = Testa thickness measured using digital calliper or micrometer, Sr = Testa roughness, Sd = Seed desiccation. QTLs for both Testa thickness measurements overlapped, except for TT5, and therefore considered as single QTL (TT). Testa roughness QTL on LGVI was relabelled GTY.

4.3.2.3 Timing of Flowering

Flowering time is an important adaptive feature and widely recognised as a major selective target for domestication. Flowering time adaptations commonly involve the relaxation of floral delay mechanisms manifesting in earlier flowering under non-inductive photoperiods (Weller et al., 2012) or in the absence of vernalisation conditions (Nelson et al., 2017). Such changes have enabled shorter growth cycles and permitted an expansion in geographical range in many crops. To compare variation in flowering time habits, two different measurements were recorded; first flowering node (*FN*) and days to flower (*DTF*).

As shown in Figure 4.7, five DTF and four *FN* QTLs were detected, which showed tight co-localisation, with only *DTF3b* not associated with a *FN* QTL. In each case the corresponding, *DTF* and *FN* QTLs mapped <8cM apart, consistent with their control by the same underlying gene. These have subsequently been relabelled as *DTF* QTLs. In total five independent QTLs were identified, consisting of one major QTL (*DTF6*) and four minor QTLs (Figure 4.7). The positions of *DTF3a* and *DTF6* were consistent with their identity as the *HR/ELF3* and *QTL6* loci identified a previous study using the F2 of

the same JI1794 x NGB5839 cross (Weller et al., 2012). QTL *FT3b* mapped to the location of the *Le* gene (Weeden et al., 1998), known to be a gibberellin 3B-hydroxylase gene (*GA3ox1*); a major regulator of gibberellin biosynthesis and modulator of plant growth and development (Lester et al., 1997, Reinecke et al., 2013). Mutations in the *Le* gene have shown to effect days to flower, but not node of flower initiation (Murfet and Reid, 1987), which is also consistent with findings for *FT3b*. QTL *DTF2* was mapped near the previously described *LATE FLOWERING* (*LF*) locus on LGII and *FT5* near to the *GIGAS*/ *FTa1-FTa2-FTc* flowering time cluster on LGV (Hecht et al. 2011). These loci appear to be in similar positions to two flowering time QTL identified in Weeden (2007) but these have not been characterised at the molecular level.

Trait code	Trait measured	QTL	LG	Position	Locus	LOD	Closest anchor marker	PEV	Associated genotype
12	Days to flower	DTF2	II	108.928	3540348_1	3.59	Between A (+9.301cM) and LF (3.072cM)	6.2	B
12	Days to flower	DTF3a	III	51.618	3564019_4	6.83	HR (-0.06cM)	12.6	A
12	Days to flower	DTF3b	III	267.94	4661775_2	3.89	LE (-3.609cM)	6.8	B
12	Days to flower	DTF5	V	140.519	FTc_2	3.21	FTc (0.00cM)	5.5	A
12	Days to flower	DTF6	VI	98.022	3548711_1	9.84	Between BFT/MTIC153/NT6083 (+4.651cM) and MLO1 (-0.04cM)	19.1	A
13	Flowering node	FN2/DTF2	II	112.305	4657639_1	5.42	LF (-0.305cM)	9.4	B
13	Flowering node	FN3/DTF3a	III	51.618	3564019_4	5.38	HR (-0.03cM)	9.3	A
13	Flowering node	FN5/DTF5	V	148.533	3562879_3	3.72	FTa1 (-4.707cM)	6.3	A
13	Flowering node	FN6/DTF6	VI	98.564	FTa3_1	15.04	FTa3 (0.00cM)	31	A

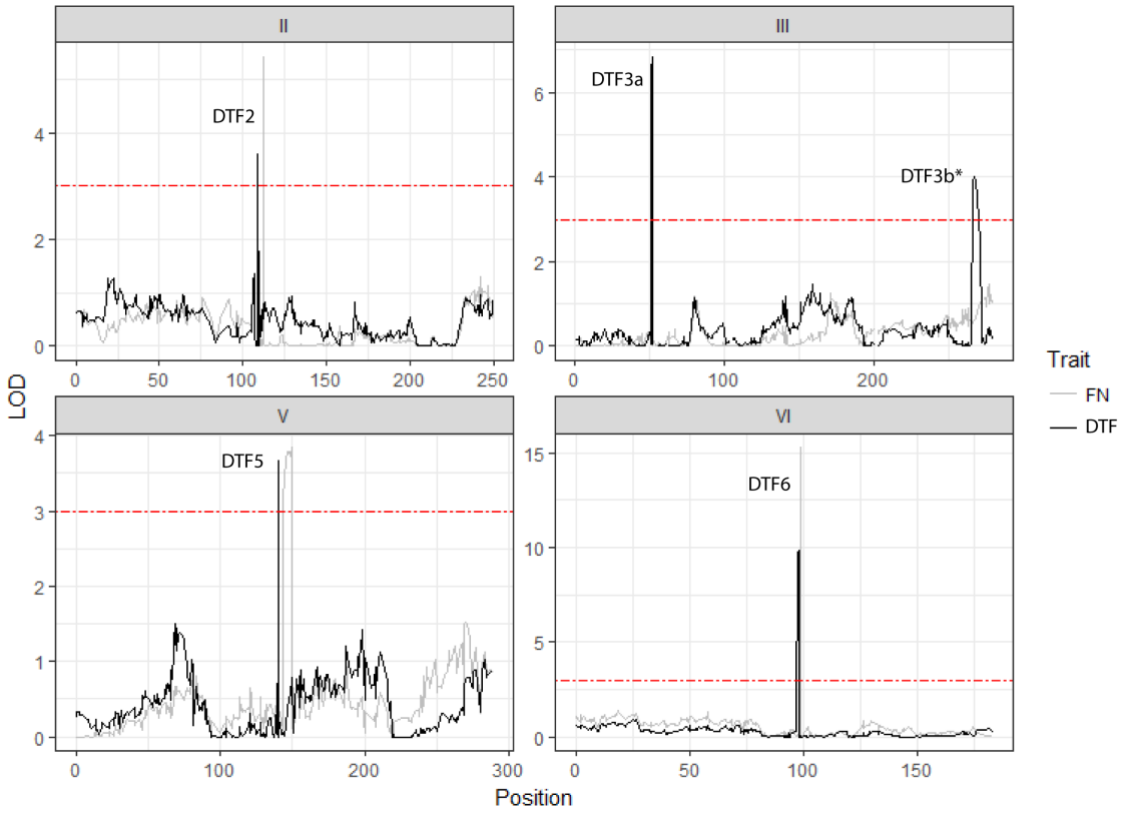


Figure 4. 7: QTL analysis results for Flowering time related traits a) Table detailing position, effect and allelic direction of all QTLs identified b) visual representation of all QTLs identified; DTF = days to flower, FN = Node to flowering. All novel flowering time QTLs were called *FT*. Only *FT3b* did not show corresponding FN QTL.

4.3.2.4 Growth habit and plant architecture

The domestication syndrome is commonly associated with a more robust, compact and determinate growth habit, reduced branching and stronger apical dominance (Hammer, 1984, Gepts, 2010). These changes are all proposed to promote productivity, reduce competition with neighbouring plants (Donald, 1968) and increase harvest efficiency (Zeder, 2015). To identify loci related to variation in these features, a total of 15 traits were measured, roughly classified into seven trait groups; i) Plant height (*Phi* and *Phii*), ii) Branching (*Bri*, and *Brii*), iii) Internode length (*INL*, *INM* and *INT*), iv) Stem thickness (*Ust* and *Lst*), v) Petiole length (*Pli* and *Plii*) vi) Pod architecture (*PPA*, *SPA* and *DPA*) and vii) Probability of dual pods (*DPP*) (see Table 4.1 for details). It is recognised some of these traits may not have arisen because of domestication but may instead represent subsequent adaptation, nevertheless they provide an overview of what are fairly typical differences in plant architecture between wild and domesticated peas.

As shown in Figure 4.8, the total number of QTLs per trait ranged from one (*Bri*) to six (*PLi*). Almost all traits were controlled by at least one major QTL, except for *DPP, Lst* and *SPA*, which also had relatively low PEV scores of 21.6%, 45.7% and 20.4% respectively. There was a strong co-localisation of QTLs to

three genomic regions which included two on LGIII, near *Hr* and *Le* respectively, and one on LGVI near the *DTF6*.

As expected, a major genetic component of plant height (*Phii*) was strongly associated with the *Le* locus. However, additional minor QTLs also mapped near flowering time regions *DTF3a* and *DTF6*, which is consistent to findings in other crop species (Durand et al., 2012, Wang et al., 2014, Zhou et al., 2016, Shen et al., 2018), while *Phii7* on LGVII mapped in the vicinity of PsCRY which encodes a DELLA protein involved in the gibberellin signalling pathway (Weston et al., 2008). Also, as expected, these loci aligned closely with the QTL for internode length, apart from *Phii7*.

QTL for relative basal branching (*Brii*) showed strong colocation with those for with plant height. This is consistent with previous findings showing strong correlation between these traits (Zheng et al., 2017, Shen et al., 2018) and may indicate possible pleiotropic effects of underlying genes on both traits. In the previous study by Weeden (2007) a single QTL was identified which was hypothesised to be *RAMOSUS1* (*RMS1*) region based on its proximity and known affect (Beveridge, 2000). As shown in Figure 4.4, *Brii3a* also mapped near *RMS1*, however we considered this QTL to be more likely due to the *ELF3* gene. This is because a loss of function mutation for *ELF3* is known to be segregating within the population (Weller et al. 2012) and has previously been shown to influence branching habit (Lejeune-Hénaut et al., 2008, Rameau et al., 2014). *Brii7* mapped near the branching gene *RMS4* on LGVII, which could be considered a strong candidate. This was supported by mapping *RMS4* and *Brii7* peak markers to *M. truncatula* reference genome which showed strong linkage (*RMS4* = Medtr4g080020, *Brii7* = Medtr4g07779).

Trait code	Trait measured	QTL	LG	Position	Locus	Anchor	LOD	PEV	Associated genotype
14	Plant height 22 days	Phi3	III	262.876	3540062_4	LE (+1.455cM)	44.89	74.5	A
14	Plant height 22 days	Phi6	VI	68.63	4663013_4	FULa (-9.452cM)	6.23	4.9	B
15	total plant height	Phii3a	III	51.846	4662891_3	HR (-0.258cM)	6.58	5.5	A
15	total plant height	Phii3b	III	262.876	4661529_2	LE (+1.455cM)	40.76	66.5	A
15	total plant height	Phii6	VI	69.212	3546994_1	FULa (-8.87cM)	3.72	3.0	B
15	total plant height	Phii7	VII	123.785	3566504_3	DUF (-16.771cM)	4.27	3.5	B
16	relative branch 22 days	Bri3	III	267.94	4661775_2	LE (-3.609cM)	6.52	19.9	B
17	Total relative branch length nodes 1 and 2	Brii3a	III	51.748	3565821_2	HR (-0.16cM)	6.24	5.0	A
17	Total relative branch length nodes 1 and 2	Brii3b	III	262.876	3540062_4	LE (+1.455cM)	41.32	65.6	A
17	Total relative branch length nodes 1 and 2	Brii6	VI	68.63	3568462_2	FULa (-9.452cM)	4.03	3.1	B
17	Total relative branch length nodes 1 and 2	Brii7	VII	123.785	3566504_3	DUF (-16.771cM)	4.21	3.3	B
18	IN 3 to 6 (internode length 3-6)	INL3	III	262.876	4661529_2	LE (+1.455cM)	37.94	63.3	A
18	IN 3 to 6 (internode length 3-6)	INL6	VI	68.557	3541050_3	FULa (+9.525cM)	11.12	11.0	B
19	IN 6 to 9 (internode length 6-9)	INM3a	III	51.748	3565821_2	HR (-0.16cM)	3.41	2.3	B
19	IN 6 to 9 (internode length 6-9)	INM3b	III	262.876	3540062_4	LE (+1.455cM)	45.63	69.4	A
19	IN 6 to 9 (internode length 6-9)	INM6	VI	68.557	3541050_3	FULa (+9.525cM	12.55	9.9	B
20	Internode length top 3 nodes	INT3	III	262.876	3540062_4	LE (+1.455cM)	17.18	38.1	A
20	Internode length top 3 nodes	INT6	VI	68.713	3547906_2	FULa (+9.525cM	7.28	13.5	B
21	Lower stem thickness	Lst3a	III	43.191	353889_2	HR (+8.397cM)	3.58	6.9	B
21	Lower stem thickness	Lst3b	III	262.876	3540062_4	LE (+1.455cM)	6.68	13.6	B

21	Lower stem thickness	Lst5	V	167.268	4662024_1	FTa1 (-24cM)	6.81	13.9	B
21	Lower stem thickness	Lst6	VI	93.371	BFT_2	BFT (0cM)	5.63	11.3	B
22	upper stem thickness	Ust3	III	50.644	3550667_2	HR (+0.944cM)	8.99	20.7	A
22	upper stem thickness	Ust5	V	93.881	3555661_2	-	3.19	6.6	B
22	upper stem thickness	Ust6	VI	69.212	3546994_1	FULa (+8.87cM)	6.96	15.4	B
22	upper stem thickness	Ust7	VII	67.265	3564090_3	-	3.54	7.3	B
23	Petiole length at node 7	PLi3a	III	51.846	4662891_3	HR (-0.258cM)	5.73	3.7	B
23	Petiole length at node 7	PLi3b	III	267.94	4661775_2	LE (-3.609cM)	29.34	28.9	A
23	Petiole length at node 7	PLi5	V	159.368	3551302_3	FTa1 (-15.542cM)	14.78	11.0	B
23	Petiole length at node 7	PLi6a	VI	72.046	4655251_1	FULa (+6.036cM)	7.31	4.8	B
23	Petiole length at node 7	PLi6b	VI	94.145	4663634_4	Between BFT/MTIC153/NT6083 (+0.774cM) and MLO1 (-3.917cM)	4.38	2.7	B
23	Petiole length at node 7	PLi6c	VI	119.959	3554231_3	BRC2 (+0.538cM)	4.48	2.8	B
24	Petiole length at node 13	PLii3a	III	51.748	3565821_2	HR (-0.16cM)	5.24	7.3	A
24	Petiole length at node 13	PLii3b	III	265.786	3560840_2	LE (-1.455cM)	18.72	33.4	A
24	Petiole length at node 13	PLii6	VI	69.212	3546994_1	FULa (+8.87cM)	10.88	16.8	B
24	Petiole length at node 13	PLii7	VII	58.832	4657109_3	DUF (48.182cM)	5.78	8.1	B
25	Primary pod axis	PPA3a	III	51.748	3565821_2	HR (-0.16cM)	11.14	18.0	A
25	Primary pod axis	PPA3b	III	262.876	4661529_2	LE (+1.455cM)	15.58	27.5	A
25	Primary pod axis	PPA6	VI	69.212	3551697_1	FULa (+8.87cM)	7	10.4	B
25	Primary pod axis	PPA7	VII	96.346	3544432_3	DUF (-10.688cM)	5.14	7.4	B
26	Secondary pod axis	SPA1	I	37.937	4663067_1		4.75	4	B
26	Secondary pod axis	SPA2	II	223.742	3553927_1		4.98	4.3	B
26	Secondary pod axis	SPA3a	III	51.161	3549664_1	HR (-0.427cM)	8.76	8.0	A
26	Secondary pod axis	SPA3b	III	110.087	3544976_1		4.86	4.1	B
27	Dual pod axis	DPA3	III	228.417	3555306_3		4.19	34.3	A
28	Dual pod probability	DPP3	III	51.618	3564019_4	HR (-0.03cM)	3.45	10.4	A
28	Dual pod probability	DPP6	VI	125.405	3543110_3	BRC2 (-4.908cM)	3.7	11.2	B

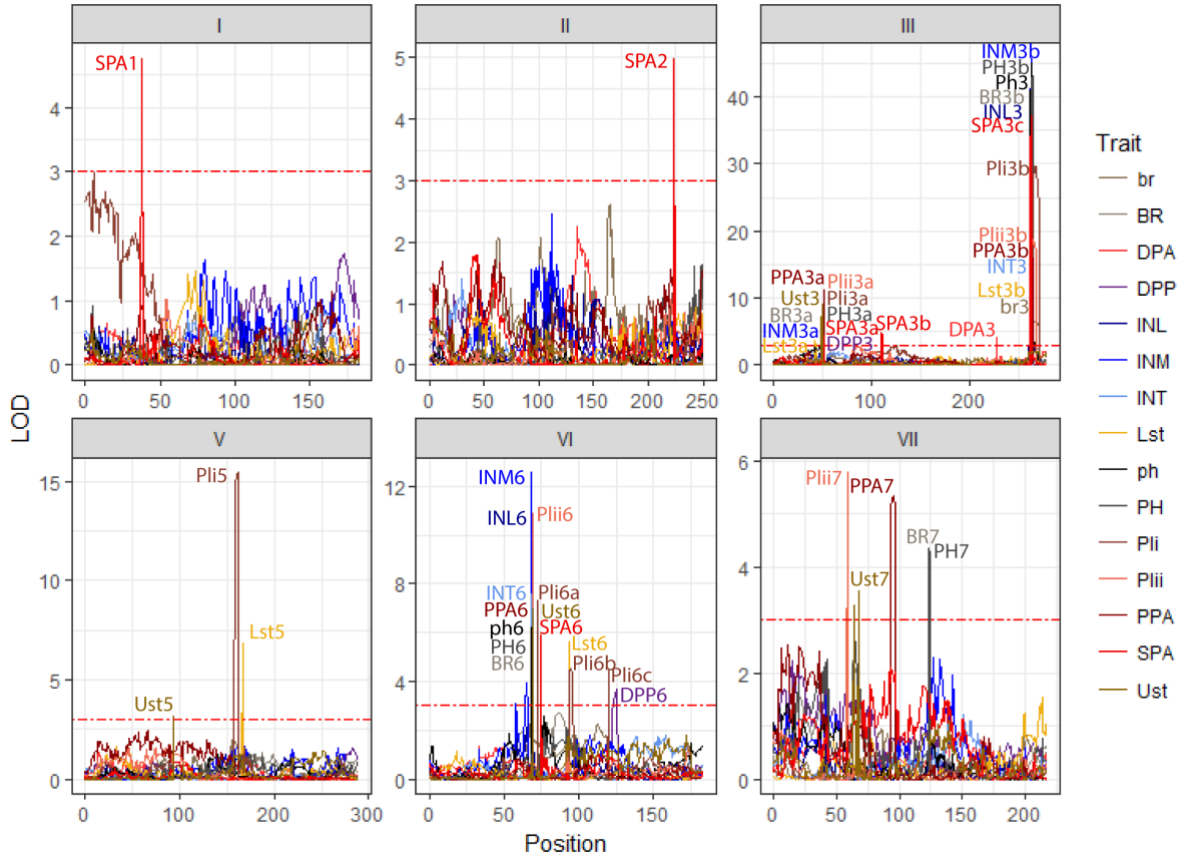


Figure 4. 8: QTL analysis results for plant architecture related traits a) Table detailing position, effect and allelic direction of all QTLs identified b) visual representation of all QTLs identified; Plant height at 22 days (Phi) and end of life cycle (Phii), branching at 22 days (Bri) and end of life cycle (Brii), internode length between nodes 3 and 6 (INL) 6 and 9 (INM) and top three node (INT), lower (Lst) and upper (Ust) stem thickness, petiole length at node 7 (Pli) and 13 (Plii), length of primary (PPA) secondary (SPA) and dual (DPA) pod axis and probability of dual pods (DPP).

4.3.2.5 Productivity

Domesticated crops tend to be enlarged relative to wild progenitors, a phenomenon often referred to as gigantism. This is often apparent in both vegetative and reproductive organs, including edible parts such as fruits and seeds. As a direct measure of productivity seed weight (*Sw*) and seed number (*Sn*) were analysed. Potentially related traits were also evaluated, which included plant growth duration (*Gd*), number of floral nodes (*Rn*) and Neoplasticity (*Np*). Neoplasticity (*Np*) is an undifferentiated non-meristematic cell growth occurring within the pod surface. This is associated with certain wild pea lines and has been interpreted as a defence mechanism against weevil oviposition (Doss et al., 1995, Teshome et al., 2016). Although not strictly recognised as a domestication-related trait, *Np* is generally not seen in domesticated lines (including NGB5839) and has been associated with reduced seed size (Weeden, 2007).

As shown in Figure 4.9, at least one major QTL was identified for each trait, with varying numbers of minor QTLs ranging between two (*Sw*) and five (*Sn*). Significant co-localisation of QTLs was identified around *Hr* and *Le* Loci. A major neoplasticity QTL (*Np3b*) mapped to the expected location of *Np* on LGIII consist with previous reports of close linkage with *Le* (Murfet and Ellis, 1998, Weeden et al., 1998, Abbo and Gopher, 2017); in addition a second novel minor QTL was observed near the *Hr* locus.

As expected, plants carrying domesticated alleles of relevant QTL generally had a larger seed size (*Sw*), but a lower seed number (*Sn*). This observation is consistent with Sadras (2007) and is likely caused by trade-offs in finite resources. When studying this at the molecular level, the major *Sw* and *Sn* QTLs (*SW3* and *Sn3c*) co-located near *Np* and *Le* loci, respectively. This is consistent with previous reports of loss of neoplasticity causing larger seed size (Weeden, 2007).

Trait code	Trait measured	QTL	LG	Position	Locus	Closest anchor marker	LOD	PEV	Associated genotype
29	Seed weight	Sw2	II	98.906	3553850_3	A (+0.721cM)	4.5	10.3	A
29	Seed weight	Sw3	III	242.887	3566597_1	LE (+21.444cM)	13.04	35.1	B
30	Seed number	Sn3a	III	51.161	3549664_1	HR (-0.427cM)	4.95	8.2	A
30	Seed number	Sn3b	III	232.6	5251779_3		5.53	9.3	B
30	Seed number	Sn3c	III	262.876	3540062_4	LE (+1.455cM)	11.35	21.2	A
30	Seed number	Sn5	V	175.341	4662522_3	FTa1 (+19.035cM)	5.18	8.7	A
30	Seed number	Sn7	VII	166.851	3551272_2		4.64	7.7	B
31	Growing duration	Gd3	III	52.532	4663129_3	HR (-0.944cM)	15.06	30.7	A
31	Growing duration	Gd5	V	150.066	3545005_1	FTa1 (-6.24cM)	4.43	7.5	A
31	Growing duration	Gd6	VI	98.061	MLO1_1	MLO1_1 (0.00cM)	6.72	11.8	A
31	Growing duration	Gd7	VII	64.148	3547099_3		3.16	5.2	B
32	Number of Flowering nodes	Rn3a	III	36.515	3546063_2	HR (+15.73cM)	3.57	6.2	B
32	Number of Flowering nodes	Rn3b	III	51.161	3549664_1	HR (-0.427cM)	9.8	18.9	A
32	Number of Flowering nodes	Rn3c	III	262.876	3540062_4	LE (+1.455cM)	9.2	17.5	A
32	Number of Flowering nodes	Rn7	VII	97.384	3563132_4	DUF (+9.63cM)	5.76	10.3	B
33	Neoplasticity	Np3a	III	67.389	3556691_1	HR (-15.801cM)	5.75	5.9	B
33	Neoplasticity	Np3b	III	248.577	3554026_1	LE (+15.754cM)	37.17	69.4	A

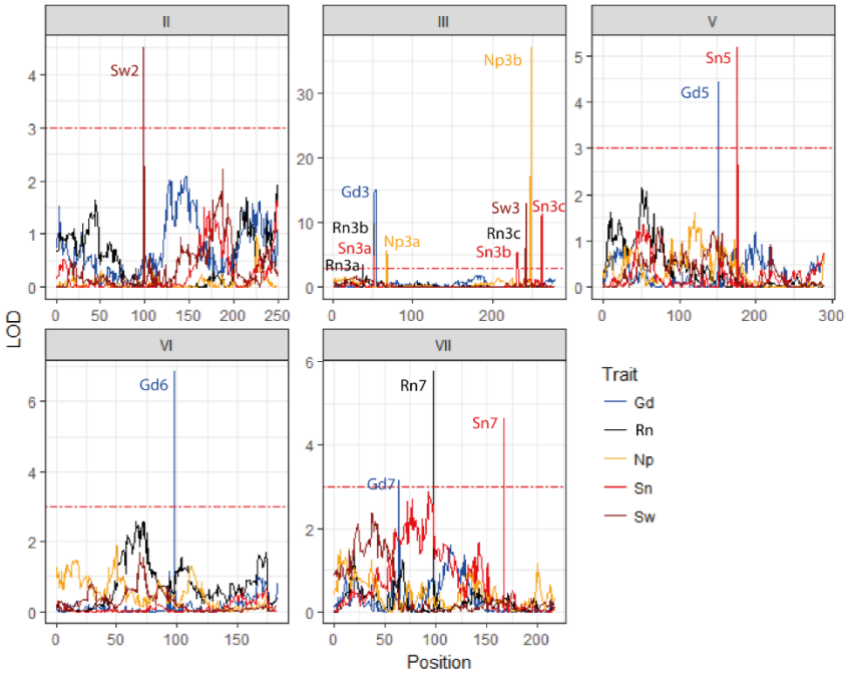


Figure 4. 9: QTL analysis results for productivity related traits a) Table detailing position, effect and allelic direction of all QTLs identified b) visual representation of all QTLs identified; Sw = Seed weight, Sn = number of seeds, Gd = growth duration, Rn = Number of floral nodes on main stem, Np = Neoplasticity.

4.3.2.6 Development rate

In response to selective pressures for shorter growth cycles and maturation rates, it is possible that domesticated forms might show an acceleration of vegetative growth, in addition to the changes in flowering time and node described earlier. To explore this possibility, seven traits were measured. These included a measure of the timing of increase in leaf complexity (node change to four leaflets (*Nc*)) and several measures of the rate of node expansion,

In pea the number of leaflets on each compound leaf increases with age and has been considered as a possible indicator of vegetative phase change (Barber, 1959, Gould et al., 1992, Wiltshire et al.,

1994). To measure development rate over time, the number of nodes expanded was measured between; i) 7 to 13 days (*Dri*), ii) 34 to 41 days (*Drii*) and iii) 49 to 56 days (*Driii*). To increase confidence this was cross referenced with total number of nodes expanded between i) 0 to 13 days (*Nei*), ii) 0 to 41 days (*Neii*) and iii) 0 to 56 days (*Neiii*).

As shown in Figure 4.10, there was strong correlation in the positions of *Ne* and *Dr* QTLs which gave confidence to their validity and position. Based on the total number of novel *Dr* and *Ne* QTLs identified, a total of nine developmental rate QTLs were recognised, which for clarity have been relabelled *Dev*. The three *Nc* QTLs corresponded with *Dev3a*, *Dev3b* and *Dev6a* map position and were therefore considered the same QTLs. Expression of these QTLs were not consistent over time, but instead differed depending on QTL and life stage, which could represent different genes being expressed during different growth phases (Poethig, 1990).

Strong co-localisation of QTLs was identified in five regions, this included one on LGII, two on LGIII and two on LGVI. These were predominately mapped near flowering time QTLs which included *DTF2*, *DTF3a*, *DTF3b DTF5* and *DTF6*. This is consistent with shorter life cycle and maturation rate in earlier flowering plants (Orf et al., 1999, Gaur et al., 2015). The plants carrying earlier flowering QTL alleles also showed reduced rate of node expansion, consistent with a reduced vegetative stage. However, this contrasts with the allelic direction of *Nc3a and Nc3b* which showed a delay in the transition to four leaflets in plants carrying the earlier flowering domesticated alleles. This was unexpected if leaf change is a marker of phase change, as it would indicate that earlier flowering was associated with an extended phase of vegetative juvenility.

Trait no.	Trait measured	QTL		LG	Position	Locus	Anchor	LOD	PEV	Associated genotype
34	Node to change to 4 leaflets	Nc3a	Dev3a	III	51.161	3549664_1	HR (-0.427cM)	10.55	23.7	A
34	Node to change to 4 leaflets	Nc3b	Dev3b	III	261.535	3569442_1	LE (+2.796cM)	4.03	9.9	A
34	Node to change to 4 leaflets	Nc6	Dev6a	VI	92.13	5252019_3	Between CABB (+0.44cM) and GA20ox (-0.67cM)	9.23	17.0	B
35	No. nodes expanded at 13 days	Nei2	Dev2	II	107.1	5251908_1	Between A (+7.473cM) and LF (-4.9cM)	3.73	7.1	B
35	No. nodes expanded at 13 days	Nei4	Dev4	IV	164.515	4661263_1	-	3.62	6.8	A
35	No. nodes expanded at 13 days	Nei6a	Dev6a	VI	98.836	4662908_3	Between FTa3 (+0.272cM) and AGOI (-0.203cM)	9.29	21.5	A
35	No. nodes expanded at 13 days	Nei6b	Dev6b	VI	174.985	4660960_3	-	3.84	9.1	B
36	No. nodes expanded at 41 days	Neii2	Dev2	II	107.622	3566442_4	Between A (+7.995cM) and LF (-4.378cM)	6.56	6.4	B
36	No. nodes expanded at 41 days	Neii3a	Dev3a	III	120.518	4657524_2	-	3.8	3.5	A
36	No. nodes expanded at 41 days	Neii3b	Dev3b	III	267.94	4661775_2	Le (-3.609cM)	27.76	40.4	A
36	No. nodes expanded at 41 days	Neii6a	Dev6a	VI	103.142	4663262_4	Between AGOI (+4.103cM) and COP13 (-0.206cM)	12.21	13.2	A
36	No. nodes expanded at 41 days	Neii6b	Dev6b	VI	175.085	3558401_1	-	5.46	5.2	B
37	No. nodes expanded at 56 days	Neiii1	Dev1	I	28.671	4662710_2	-	4.17	3.2	B
37	No. nodes expanded at 56 days	Neiii2	Dev2	II	112.305	4657639_1	LF (+0.305cM)	9.06	7.6	B
37	No. nodes expanded at 56 days	Neiii3a	Dev3a	III	52.532	4663129_3	HR (0.944cM)	7.87	12.4	A
37	No. nodes expanded at 56 days	Neiii3b	Dev3b	III	265.786	3560840_2	LE (1.455cM)	12.02	17.9	A
37	No. nodes expanded at 56 days	Neiii5	Dev5	V	175.341	4662522_3	FTa1 (19.035cM)	4.21	3.2	A
37	No. nodes expanded at 56 days	Neiii6a	Dev6a	VI	98.061	MLO1_1	MLO1 (0.00cM)	14.26	16.8	A
37	No. nodes expanded at 56 days	Neiii6b	Dev6b	VI	173.182	3555794_2	-	6.1	7.6	B
38	Development rate 7-13	DRi6	Dev6a	VI	101.53	3556028_4	Between AGOI (+2.491cM) and COP13 (-1.818cM)	9.01	26.5	A
39	Development rate 36 - 41 days	DRii2	Dev2	II	112	LF_2	LF (0.00cM)	3.33	7.1	B
39	Development rate 36 - 41 days	DRii3a	Dev3a	III	57.066	3541250_3	HR (5.478cM)	3.48	10.5	A
39	Development rate 36 - 41 days	DRii3b	Dev3b	III	267.94	4661775_2	Le (-3.609cM)	8.24	22.9	A
40	Development rate 49-56 days	DRiii3	Dev3	III	51.659	4663518_2	HR (0.0.71cM)	14.34	28.5	A
40	Development rate 49-56 days	DRiii5	Dev5	V	141.675	3544855_2	FTa1 (14.631cM)	4.74	7.9	A
40	Development rate 49-56 days	DRiii6	Dev6	VI	98.061	MLO1_1	MLO1	6.12	10.5	A
40	Development rate 49-56 days	DRiii7	Dev7	VII	114.47	5252065_3	DUF (7.456cM)	4.51	7.5	B

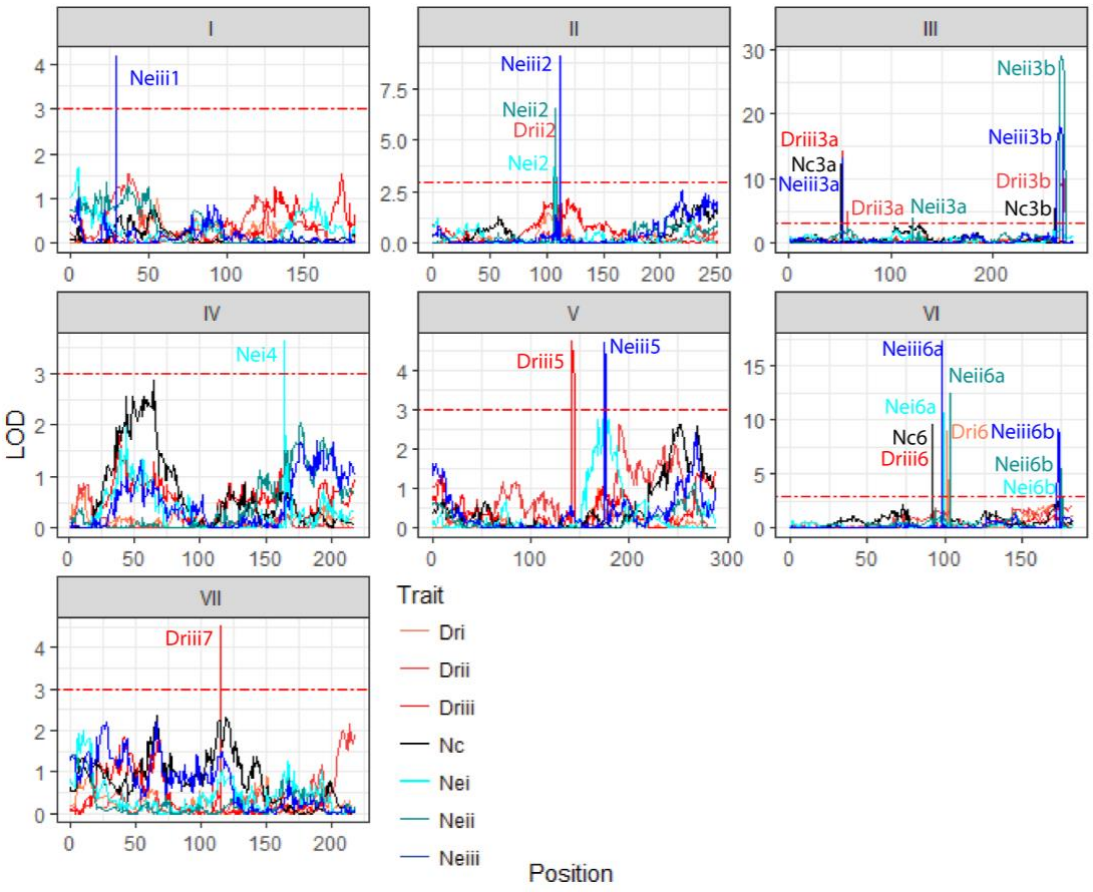


Figure 4. 10: QTL analysis results for Developmental related traits a) Table detailing position, effect and allelic direction of all QTLs identified b) visual representation of all QTLs identified; Node change (Nc), Nodes expanded between 7-13 days (Devi), 36-49 days (Devii) and 49 – 56 days (Deviii), and total number of nodes expanded at 13 days (Nei), 49 days (Neii) and 56 days (Neiii)

4.3.2.7 Leaf traits

Modification in leaf morphology are typically not been discussed as a domestication trait, but as significant differences in leaf size and shape were observed between parental lines, these traits were also assessed. Whether these variations provide a selective advantage for cultivation is unknown and was not a focus of this study. The domesticated parent NGB5839 showed an increase in leaf size and roundness and a reduced level of serrations relative to JI1794. In total, nine leaf morphology measurements were analysed, four for leaf size (*LfH*, *LfW*, LfW and *LfA*) and five for leaf shape (*HwP*, *FvT*, *Pci*, *Pcii*, *LfS*) (Table 4.1).

As shown in Figure 4.11, almost all traits gave at least one major QTL excluding *LfA* and *FvT,* with minor QTLs varying between one (*FvT*) and four (*LfS*). As expected, plants carrying the domesticated QTLs alleles had larger leaves, except for *LfH3a*, which is consistent with its location close to the *Le* gene. Findings of larger leaf size with plants carrying domesticated allele is consistent with other legume crops such as Cowpea (Lo et al., 2018), Mungbean (Isemura et al., 2012) Yardlong (Kongjaimun et al., 2012) Azuki bean (Kaga et al., 2008) and Rice bean (Isemura et al., 2010). The tendency for

domesticated crops to exhibit larger leaves suggests that it may have a selective advantage, such as increased photosynthetic potential, but these traits were not investigated in this study. For leaf shape there was no consistent allelic direction, with traits either showing low PEV score and/or relatively high number of minor QTLs. This suggests a complex genetic control with little effect of the major dwarfing gene *Le*, and that these had little or no selective advantage.

Trait code	Trait measured	QTL	LG	Position	Locus	Closest anchor marker	LOD	PEV	Associated genotype
41	Leaf length	LfH3a	III	102.675	3542109_4		5.08	8.2	B
41	Leaf length	LfH3b	III	260.086	3554067_1	LE (+4.245cM)	3.64	5.7	A
41	Leaf length	LfH5	V	159.368	3551302_3	FTa1 (-15.542cM)	4.15	6.6	B
42	Leaf length	LfH6	VI	93.574	3565779_3	Between BFT/MTIC153/NT6083 (+0.203cM) and MLO1 (-4.488cM)	15.3	29.6	B
42	Leaf width	LfW4	IV	157.578	4663045_4		6.79	16.1	B
42	Leaf width	LfW6	VI	68.713	5252250_1	FULa (+9.39cM)	7.99	19.4	B
43	Leaf perimeter	LfP5	V	81.553	3557605_3		4.07	8.4	B
43	Leaf perimeter	LfP6	VI	98.022	3548711_1	Between BFT/MTIC153/NT6083 (+4.651cM) and MLO1 (-0.04cM)	11.27	26.3	B
43	Leaf perimeter	LfP7	VII	41.586	4661178_1		3.53	7.2	B
44	leaf area	LfA4	IV	158.934	4658946_1		4.28	8.2	B
44	leaf area	LfA5	V	113.847	3543940_2	FTc (+26.672cM)	3.52	6.6	B
44	leaf area	LfA6a	VI	69.212	3551697_1	FULa (+8.87cM)	9.72	7.6	B
44	leaf area	LfA6b	VI	107.363	3554624_2	Between COP13 (+4.015cM) and BRC2 (-13.134cM	9.21	6.5	B
45	height widest point	HwP1	I	88.37	4663570_1	-	3.18	5.5	B
45	height widest point	HwP3	III	271.452	4660684_3	LE (-7.121cM)	6.34	11.7	A
45	height widest point	HwP6	VI	98.836	4662908_3	Between FTa3 (+0.774cM) and APRL (-0.203cM)	15.52	33.8	B
46	Leaf roundness	FvT4	IV	143.334	3557818_3	-	3.97	12.7	A
47	PC1	Pci2	II	96.5	3564794_1	Mendel A (-3.127cM)	9.33	14.8	B
47	PC1	Pci3	III	46.32	4662321_2	HR (+5.268cM)	5.74	8.5	A
47	PC1	Pci4	IV	122.957	4663001_1	-	6.29	9.4	B
47	PC1	Pci6a	VI	68.713	3547906_2	FULa (9.369cM)	3.69	5.3	B
47	PC1	Pci6b	VI	93.955	3542173_1	Between BFT/MTIC153/NT6083 (+0.584cM) and MLO1 (-4.107cM	15.64	27.8	A
48	PC2	Pcii3a	III	27.763	3565032_1	HR (23.825cM)	4.24	5.9	B
48	PC2	Pcii3b	III	262.279	4662578_2	LE (+2.052cM)	3.86	5.4	B
48	PC2	Pcii5	V	30.662	4661973_3	-	3.47	4.8	B
48	PC2	Pcii6a	VI	74.661	3534742_1	FULa (+3.421cM)	15.34	26.2	A
48	PC2	Pcii6b	VI	126.95	3565911_3	BRC2 (-6.453cM)	4.72	6.6	A
49	Serrations	LfS1	I	77.209	3559103_3	-	15.85	24.2	A
49	Serrations	LfS2	II	111.738	3566713_1	LF (+0.262cM)	3.27	4.0	A
49	Serrations	LfS3	III	271.452	4660684_3	LE (-7.121cM)	11.73	16.6	B
49	Serrations	LfS4	IV	193.774	3546560_2	-	5.87	7.5	B
49	Serrations	LfS6a	VI	9.806	3543628_4	-	3.53	4.3	A
49	Serrations	LfS6b	VI	62.928	3551782_1	FULa (+15.154cM)	8.45	11.3	B

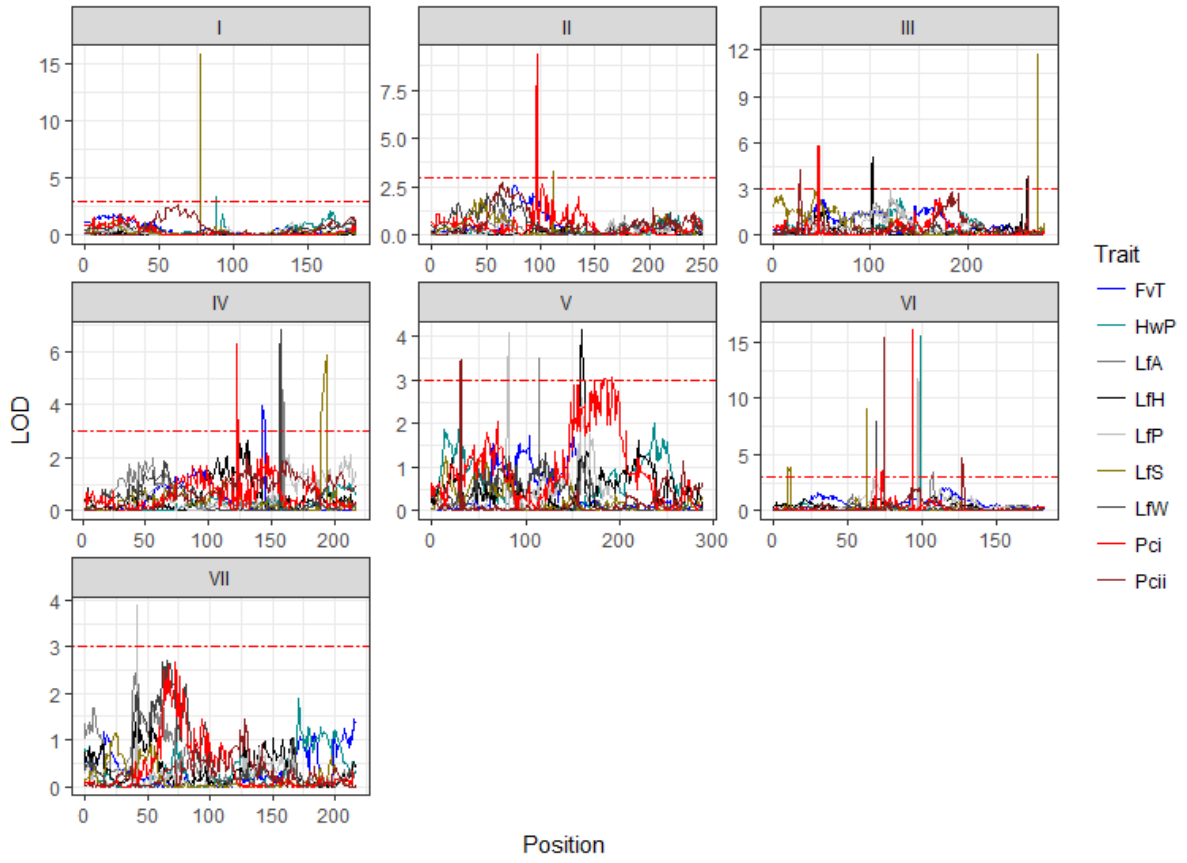


Figure 4. 11: QTL analysis results for leaf morphology related traits a) Table detailing position, effect and allelic direction of all QTLs identified b) visual representation of all QTLs identified; Leaf length (LfH), Leaf width (LfH), Leaf perimeter (LfP), and Leaf area LfA as well as leaf shape: Height at widest point (HwP), Leaf roundness (FvT), leaf shape principle component 1 (Pci) and 2 (Pcii) and leaf serrations (LfS)

4.3.3 Clustering of domestication related traits

As shown in Figure 4.4, QTLs were not evenly distributed, but instead showed a high a concentration to four genomic regions. These regions were referred to as Clst2, Clst3a, Clst3b and Clst6 which relate to LGII between 92.6cM and 112.305cM; LGIII between 27.763 and 67.389cM; LGIII between 250.090cM and 271.452cM and; LGVI between 62.928cM and 107.363cM respectively. As shown in Table 4.5, 102 (17+25+25+35 = 102) of the total 160 QTLs were found within these four regions, with 15 of the 49 traits having QTLs mapping to three or more of the four clusters. Collectively this suggests that many of these traits could be pleiotropic consequences of variation in a single causal gene. In particular, flowering time pathways are known to interact with a diverse range of other developmental and growth habit pathways and as flowering time QTLs mapped to all four QTL clusters these could potentially be regulating many of the other traits.

Table 4. 5: Traits identified in within clustered regions; Clst2, Clst3a, Clst3b and Clst6. Numbers refer to trait code, highlighted regions denote where same trait is identified in 3 or more of the clustering regions

Trait	Clst2	Clst3a	Clst3b	Clst6
Total flower colour	X			
Green testa	X			
Red testa	X	X		
Mottled	X	X		
Black spots	X			
Black hilum				
Permeability	X			
Testa thickness Digital Calipers	X			X
Testa thickness - Micrometer	X			X
Testa roughness				X
Seed desiccation			X	
Flowering time	X	X	X	X
Floral node	X	X		X
Plant height at 22 days			X	X
Total plant Height		X	X	X
Relative branch length at 22 days			X	
Total relative branch length nodes 1 and 2		X	X	X
Internode length between 3 and 6			X	X
Internode length between 6 and 9		X		X
Internode length between top 3 nodes of plant			X	X
Lower stem thickness		X	X	X
Upper stem thickness		X		X
Length of petiole at node 7		X	X	X
Length of petiole at node 13		X	X	X
Primary pod axis		X	X	X
Secondary pod axis		X		
Dual pod axis				
Probability of dual pods		X		
Seed weight	X			
Seed number		X	X	
Duration of plant growth		X		X
Number of floral nodes		X	X	
Neoplasticity		X		
Node to change to 4 leaflets		X	X	X
Number of nodes expanded at 13 days	X			X
Development rate between 7-13 days	X		X	X
Number of nodes expanded at 41 days	X	X	X	X
Development rate between 36 and 41 days				X
Number of nodes expanded at 56 days	X	X	X	
Development rate between 49 and 56 days		X		X
Leaf length			X	
Leaf width				X
Relative height of leaf at widest point				X
Leaf roundness				X
Leaf perimeter			X	X
Leaf area				
Leaf shape principal component 1	X	X		X
Leaf shape principal component 2		X	X	X
Leaf serrations	X		X	X
Total	17	25	25	35

To explore the potential pleiotropic effect of flowering time genes, a separate near-isogenic progeny was examined in which the wild-type *Hr* allele had been introgressed through five backcrosses into the NGB5839 (*hr*) genetic background, This progeny was shown to segregate in a ratio of 6:5:5 for plants carrying wild (A): heterozygous (H): domesticated (B) alleles, using a molecular marker for the *ELF3a*/*Hr* gene, and a total of 29 traits were measured.

As shown in Table 4.6 and Supplementary Figure 4.4, 16 out of the 29 measured traits were significantly affected by the *Hr* allele, consistent with the conclusion that the segregation at *Hr* in the RIL population might indeed exert pleiotropic effects on other traits and explain at least in part the clustering of QTLs in the Clst3a region. As expected segregation of *Hr* in the NIL progeny was associated with a significant effect on development rate, with plants carrying the domesticated allele developing nodes at a faster rate than the wild, but with reduced number of total nodes. The reduced growth period in domesticated lines is likely caused by an overall acceleration of development rate, both in terms of an accelerated progression of vegetative development (earlier transition to ≥ 8 and 11 leaflets/tendrils per leaf) and earlier flowering (Wiltshire et al., 1994). Secondary effects of the domesticated *Hr* allele might include other differences common to both populations, including reduced plant height and internode length and branching, increased stem thickness, an increased number of flowering nodes on main stem and increased number of total pods and seeds.

Table 4. 6: A Tukeys pairwise comparison between different genotypes (H-B = heterozygous and domesticated, A-B = wild and domesticated and A-H = wild and heterozygous) on traits measured on a NIL segregating for HR. Values highlighted in orange signify significant differences (P ⩽ 0.05).

	Pairwise comparison		
	H-B	A-B	A-H
Plant height	0.000	0.000	0.157
Internode length nodes 3 to 6	0.715	0.022	0.096
Internode length nodes 6 to 9	0.979	0.846	0.934
Internode length 3 to 9	0.909	0.331	0.561
Petiole length at node 7	0.759	0.139	0.356
Petiole length at node 13	0.044	0.012	0.793
Flowering node	0.000	0.000	0.000
number floral nodes	0.077	0.018	0.770
Number of seeds on main stem	0.811	0.392	0.715
Number of seeds on branch	0.826	0.072	0.203
Total number of seeds	0.215	0.102	0.925
Number of pods on branch	0.782	0.084	0.263
Number of pods on main stem	0.187	0.045	0.740
Total number of pods	0.207	0.005	0.165
Branch length	0.824	0.119	0.314
Primary pod axis	0.018	0.004	0.619
Dual pod axis	0.621	0.198	0.681
Secondary pod axis	0.971	0.409	0.502
Growth duration	0.000	0.000	0.154
Lower stem thickness	0.494	0.016	0.138
Upper stem thickness	0.356	0.246	0.979
Total number of nodes	0.000	0.000	0.387
Node change to 4 or greater leaflets and tendrils	0.489	0.439	0.988
Node change to 6 or greater leaflets and tendrils	0.818	0.617	0.946
Node change to 8 or greater leaflets and tendrils	0.661	0.011	0.061
Node change to 11 or greater leaflets and tendrils	0.002	0.000	0.827
Node change to 13 or greater leaflets and tendrils	0.000	0.000	0.055
Development rate week 4 to 5 (total number of nodes)	0.385	0.040	0.402
Development rate week 8 to 9 (total number of nodes)	0.000	0.000	0.978

To provide an additional validation of the clustering of QTLs, 26 traits were phenotyped in three F4 progenies from the JI1794 x NGB5839 cross that were determined to be segregating only for *DTF6* and not for *DTF2*, *DTF3a* or *DTF3b* QTLs by molecular genotyping for *FTa3*, *LF*, *Hr* and *Le* genes respectively. Population 1, 2 and 3 consisted of 47, 24 and 28 individuals and were genotyped to show segregation of the Clst6 region with A:H:B genotype ratios of 9:27:11, 9:10:5 and 8:13:7, respectively.

As shown in Table 4.7 and Supplementary Figure 4.5, 19 of the 26 traits measured showed a significant difference between lines carrying wild and domesticated alleles in one or more of the three populations. As in the case of the *HR* NILs, it is important to state that this analysis does not prove a causal link between the flowering time gene/locus and the other traits, and the possibility that these traits may be controlled by closely linked but independent genes cannot be ruled out. However, the fact that association between flowering time and other traits was observed for two loci in two separate progenies does add support to the idea that the flowering time loci in these regions might be having pleiotropic effects. Consistent to the RILs, plants carrying the earlier flowering domesticated

genotype tended to have the opposite allelic effect on putative pleiotropic traits than the earlier flowering domesticated *Hr* genotype.

Table 4. 7: A Tukeys pairwise comparison between different genotypes (H-B = heterozygous and domesticated, A-B = wild and domesticated and A-H = wild and heterozygous) on traits measured on a 3 F4 populations segregating for FTa3 but fixed for Clst2 (Mendel A and LF), Clst3a (Hr) and Clst3b (Le). Values highlighted in orange signify significant differences (P≤ 0.05).

	Pop1			Pop2			Pop3		
	A-H	B-A	B-H	A-H	B-A	B-H	A-H	B-A	B-H
Plant height	0.043	0.023	0.713	0.571	0.745	0.994	0.064	0.437	0.633
Relative branch length	0.077	0.032	0.647	0.406	0.990	0.609	0.124	0.200	1.000
Internode length 3 to 6	0.003	0.000	0.305	0.001	0.000	0.093	0.003	0.000	0.237
Internode length 6 to 9	0.196	0.010	0.138	0.001	0.000	0.013	0.014	0.015	0.910
Internode top 3	0.075	0.084	0.905	0.064	0.274	0.949	0.098	0.001	0.034
Number of seeds	0.933	0.328	0.333	0.991	0.921	0.872	0.905	0.535	0.264
Flowering time	0.037	0.005	0.322	0.121	0.003	0.102	0.314	0.021	0.218
Flowering Node	0.003	0.002	0.628	0.000	0.000	0.002	0.004	0.000	0.001
primary pod axis length	0.280	0.083	0.445	0.338	0.292	0.931	0.301	0.825	0.701
dual pod axis length	0.868	0.376	0.379	0.487	0.398	0.883	.	.	.
Total nodes	0.000	0.000	0.647	0.008	0.000	0.034	0.341	0.014	0.128
number of nodes against plant height	0.000	0.000	0.499	0.003	0.000	0.242	0.001	0.002	0.972
lower stem thickness	0.859	0.896	0.503	0.361	0.002	0.019	0.608	0.197	0.558
Upper stem thickness	0.995	0.247	0.099	0.059	0.052	0.872	0.853	0.227	0.344
growth duration	0.000	0.000	0.274	0.065	0.001	0.095	0.880	0.185	0.289
Node when Sum of Leaf and tendril ≥ 4	0.000	0.000	0.326	0.008	0.005	0.722	0.000	0.001	0.896
Node when Sum of Leaf and tendril ≥ 6	0.000	0.000	0.001	0.326	0.003	0.037	0.000	0.000	0.008
Node when Sum of Leaf and tendril ≥ 8	0.000	0.000	0.000	0.013	0.000	0.044	0.001	0.000	0.000
Node when Sum of Leaf and tendril ≥ 10	0.000	0.000	0.078	0.056	0.034	0.779	0.508	0.060	0.274
Petiole length at node 7	0.006	0.000	0.005	0.105	0.065	0.802	0.539	0.388	0.896
Petiole length at node 13	0.004	0.002	0.589	0.559	0.935	0.864	0.854	0.587	0.830
Flowering nodes on main stem	0.805	0.205	0.301	0.574	0.544	0.971	0.627	0.990	0.693
nodes expanded at week 5	0.006	0.000	0.213	0.012	0.000	0.025	0.585	0.040	0.159
nodes expanded at week 9	0.001	0.000	0.281	0.007	0.000	0.030	0.901	0.018	0.023
Development rate between week 4 and 5	0.479	0.309	0.817	0.945	0.627	0.778	0.989	0.410	0.278
Development rate between week 8 and 9	0.805	0.205	0.301	0.574	0.544	0.971	0.627	0.990	0.693

4.4 Discussion

During domestication of pea, human selection for desirable traits has been accompanied by distinct morphological changes which distinguish domesticated forms from their wild progenitors. To investigate the underlying genetic changes this study has genetically dissected some of the changes which have occurred between wild and modern cultivar JI1794 x NGB5839. This has involved QTL analysis of 49 independent measurements using the high-density linkage map previously developed and described in chapter 3, making this one of the most comprehensive examinations of morphological variations incurred between wild and domesticated pea to date. While this chapter has given a broad overview on the genetic architecture of many different traits, the two main traits of interest for subsequent focus are seed dormancy and flowering time.

4.4.1 Genetic control of known segregating loci

From the known segregating loci (Table 4.3), strong co-linearity of *A*, *M*, *Np*, *Le*, *Fs*, *Gty* and *Pl* were observed to the Weeden et al. (1998) consensus map, and *Hr* and *QTL6* with Weller et al. (2012);

thereby giving further confidence in the quality of the map developed in Chapter 3. This is most clearly shown when comparing linkage to other known loci, as demonstrated with Mendel *A* which was mapped above but closely linked to *LF* by a distance of around 12cM, findings consistent with previous studies (Uzhintseva and Sidorova, 1988, Kosterin, 1994, Weeden et al., 1998, Hellens et al., 2010). Close linkage was also reported between *Le* and *Np* (labelled *Np3b*) by a distance of around 16cM, findings also consistent to previous reports (Murfet and Ellis, 1998, Weeden et al., 1998, Smirnova, 2002), this is not to be confused with a second minor novel *Np* QTL (labelled *Np3a*), which was identified in this study on LGIII near *Hr* (~16cM).

Whilst previous studies have mapped these loci detailed in Table 4.3 to varying levels of resolution, only *A*, *ELF3*/*Hr* and *Le* have been characterised. This study has mapped many of these loci at greater resolution than previous reports which can be used as a platform for further fine mapping in later studies.

4.4.2 Genetic control of domestication related traits

Previous molecular studies have shown domestication related traits are predominately controlled by relatively few but large-effect QTLs (Paterson, 2002, Ross-Ibarra, 2005, Weeden, 2007, Gepts, 2010). Results from this study are consistent with this conclusion, with 35% (56 of 160) of the QTLs identified as major (i.e. explaining >15% of the variance) and an average of approximately 3 QTLs per trait (49 traits/160 QTLs = 3.27). As shown in Supplementary Figure 4.2 a total of 45 of the 49 traits were controlled by one or two major QTLs, indicating that these traits underwent a distinct major phenotypic transformation. The collective evidence for simple genetic control of domestication-related traits supports the concept that domestication incurred a rapid evolution (Gross and Olsen, 2010, Abbo and Gopher, 2017). This is consistent with previous analyses of pea (Weeden, 2007), but in contrast to findings in sunflower which identified more complex genetic regulation with greater proportion of minor to intermediate QTLs (Wills and Burke, 2007).

4.4.3 Genetic control of seed dormancy

In pea, loss of seed dormancy has been widely viewed as a critical domestication trait (Ladizinsky, 1987, Zohary, 1989, Abbo et al., 2011), as it has been argued that it would expect to show rapid evolution and therefore simple genetic control (Abbo et al., 2014). This is consistent with our results which found two permeability QTLs, this including a major locus, *Perm2*, explaining 79.2% of total observed variation (PEV = 32.7%) and a minor QTL, *Perm7* explaining 20.8% of observed variation (PEV = 8.6%). Arguably, the large effect of *Perm2* is indicative of a fundamental role in overcoming hardseededness and therefore based on its critical function of loss of seed dormancy could be

considered a domestication gene. In comparison the smaller effect of *Perm7* suggest this may have been a secondary modification that acts to further promote germination, in an additive manner with *Perm2.* In the previous study, Weeden (2007) also found two seed dormancy QTLs, with the wild x cultivated population (MN313 x JI1794) showing a QTL mapping in a similar region to *Perm2* and possibly primitive x wild (WL808 × JI261) and cultivated x primitive (cultivated x WL808) as well, while *Perm7* to our knowledge is a novel QTL, supporting the concept that *Perm7* is a diversification QTL.

Similar findings where seed dormancy was controlled by one or two QTLs were also found in other legume species including lentil (Ladizinsky, 1985), common bean (Koinange et al., 1996) soybean (Liu et al., 2007, Sun et al., 2015) and azuki bean (Kaga et al., 2008). In contrast more complex genetic control of dormancy was reported in Mungbean with four QTLs (Isemura et al., 2012), five in rice bean (Isemura et al., 2010), and six in Yardlong bean (Kongjaimun et al., 2012). However, even in these cases, most of the effect was due to one major QTL (explaining 57.5%, 55.5% and 77.3% of the total PEV values, respectively). This is consistent with a scenario where a single major variant was initially adopted which sufficiently reduced dormancy to enable more synchronous germination, followed by subsequent, smaller to further reduce seed dormancy and improve germination consistency.

4.4.4 Genetic control of flowering time

Flowering time is a complex trait regulated by large number of genes interacting in an intricate network of signalling pathways. It is also well known to be a key agronomic trait, playing a pivotal role in crop adaptation and expansion to novel agro-ecological areas (Michael et al., 2003). In a previous genetic analysis of flowering time in a JI1794 x NGB5839 F2 progeny under SD conditions, Weller et al. (2012) described two QTLs; a major QTL on LGIII identified as the *EARLY FLOWERING3* (*ELF3*) gene and a minor QTL on LGVI referred to as *QTL6*. This study has re-examined this genetic control using a RIL population grown in LD conditions, which are considered more consistent with the prevailing conditions in its natural growing environments, than SD. Here, five significant flowering time QTLs were identified for which three (*DTF3a*, *DTF5* and *DTF6*) promoted earlier flowering in NGB5839, while the other two (*DTF2* and *DTF3b*) delayed flowering. The mapping positions of *DTF3a* and *DTF6* were consistent with those reported for *Hr*/*ELF3* and *QTL6* by Weller et al. (2012) and were therefore considered likely to be the same loci.

The *DTF3a* which mapped to the *ELF3* gene is part of the circadian clock and involved in the photoperiod pathway (McClung, 2006). As this QTL had previously been demonstrated within this population, its occurrence in LD conditions was not unexpected, particularly as previous studies have shown *elf3* mutants promote flowering in both SD and LD conditions (Liu et al., 2001, Turner et al., 2005, Bendix et al., 2015). Studies have shown convergent targeting of the *ELF3* gene in other crop

models including but not limited to rice (Matsubara et al., 2012), barley (Faure et al., 2012, Zakhrabekova et al., 2012) and legume species, chickpea (Ridge et al., 2017), pigeon pea (Varshney et al., 2017) and lentil (Weller et al., 2012), therefore highlighting its importance as a domestication target.

The *DTF6* which likely confers to the *QTL6* locus, has not yet been characterised at the molecular level, but may correspond to the known naturally occurring *E* locus (Murfet and Reid, 1987). This mapped to anchor marker *FTa3* which corresponds to the potentially novel *FT* homolog recently identified in *G. max*, *M. truncatula* and *C. arientinum* (Książkiewicz et al., 2016, Nelson et al., 2017). The *Fta3* ortholog has currently not been characterised in pea (Hecht et al., 2011), however its occurrence within all species examined within galegoid clade (Książkiewicz et al., 2016, Nelson et al., 2017) and a subsequent blast search in closely related species *L. culinaris* strongly indicates to its presence. Little is currently known of its functional role, however the *Fta* clade in legumes has previously been linked with vernalisation response (Hecht et al., 2011, Ortega, 2018). Examination of this region is being explored in a separate study and will not be discussed further here.

Other flowering time QTLs; *DTF2*, *DFT3b* and *DTF5* were not identified in the previous study of Weller et al. (2012). *DFT3b* mapped to the bottom of LGIII near the known induced *le-3* mutant (Lester et al., 1999) found within the cultivated parental line. *Le* encodes the enzyme gibberellin 3P-hydroxylase (GA3ox1) (Lester et al., 1997), an major regulator of bioactive gibberellin GA1 (Ingram et al., 1984, Reid and Ross, 2011). In arabidopsis, Gas have been shown to accelerate floral development (Hisamatsu and King, 2008) through the degradation of the DELLA transcriptional repressor (Murase et al., 2008, Yu et al., 2012) and by promoting expression of floral integrator genes *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*), *LEAFY* (*LFY*) and *FLOWERING LOCUS T* (*FT*) (Mutasa-Göttgens and Hedden, 2009). In previous studies, plants carrying the *le* mutant were later to flower than WT but flowered at the same node (Murfet and Reid, 1987, Mitchum et al., 2006, Hecht et al., 2007). This was consistent with findings from the present study which found a significant *DTF* QTL but no FN QTL in the region of *Le*. Together these observations strongly implicate *Le* as the as the causative gene.

DTF2 showed tight linkage to the LGII anchor marker *LATE FLOWERING* (*LF*) which corresponds to the *PsTFL1c* gene and is known to be an important modulator of flowering (Foucher et al., 2003). The arabidopsis homolog *TFL1* represses flowering by competing with *FT* in the apex for interaction with *FD,* thereby inhibiting expression of downstream target genes (Abe et al., 2005). Natural occurring variants of the *LF* gene are known to exist in domesticated pea germplasm (Weller et al., 1997) and a QTL was previously mapped in the vicinity of *LF* by Weeden (2007) study. Analysis of null *lf* mutants

has shown that *LF* delays flowering (Foucher et al., 2003), but interestingly the domesticated allele at *DTF2* conferred later flowering. If variation at *LF* was the basis for *DTF2*, this would imply an increased function in the domesticated relative to the wild allele, a less common situation (Meyer and Purugganan, 2013). A more detailed analysis to the genetic control of this QTL will be explored in chapter 7.

DTF5 mapped near to the *FT* homolog *Fta1-Fta2-FTc* cluster, a region known to be an important regulator of flowering. Pea *Fta1*/*gigas* mutants are late flowering and remain vegetative under some LD conditions, (Beveridge and Murfet, 1996, Hecht et al., 2011) indicating that *Fta1* may be essential for flowering in LD (Hecht et al., 2011). The same region has also been shown to affect flowering time in other legume species (Weller and Ortega, 2015, Jaudal et al., 2016, Rajandran, 2016, Ortega, 2018) but so far there has been no clear evidence for its importance in natural variation for flowering time in pea (Weller et al., 1997, Weller et al., 2012). The fact that the domesticated allele is associated with earlier flowering is again consistent with a gain-of-function variant for one of the *FT* genes in this cluster. Weeden (2007) did identify a flowering time QTL on LGV (*QTL-V*) the direction of the allelic effect was not stated, and the low resolution of the position made it difficult to assess potential candidate genes. A more detailed examination of this region will be examined in chapter 7.

4.4.5 Pleiotropism or clustering of domestication-related genes

In this study, QTLs demonstrated clustering in four genomic regions. This might reflect the action of that a single gene with pleiotropic effect or the action of numerous closely linked genes underlie these QTLs. Genomic clustering of QTLs has been reported in previous examinations of domestication related traits (Ross-Ibarra, 2005, Gepts, 2010), including in numerous legume species (Koinange et al., 1996, Isemura et al., 2007, Kaga et al., 2008, Isemura et al., 2010, Isemura et al., 2012, Kongjaimun et al., 2012, Lo et al., 2018). This was in contrast to soybean (Liu et al., 2007) and previous examination of pea (Weeden, 2007) which reported no clustering.

Among the 49 measurements analysed, it is recognised that not all were independent variables, consequently in several cases different trait measurements may have directly or indirectly measured the same underlying process, therefore potentially causing false-positive genomic clusters. For example, where there are no differences in the rate of node development or flower opening, flowering time (*FT*) and node of flowering (*FN*) are normally very closely correlated. Also, plant height (*Phii*) is a compound variable that integrates, internode length (*IN*) and the rate of node development. To validate the clusters were not an artefact of indiscrete traits, the 49 measurements were grouped into the seven morphological groups, as detailed in Table 4.1. If two or more QTLs from the same morphological group mapped within 10cM these were considered the same QTL. As shown in

Supplementary Figure 4.2, the reanalysing of QTL distribution validated the four genomic clusters, however whether these are caused by pleiotropic or tightly linked but independent genes is more debatable (Ross-Ibarra, 2005). To explore this question, we examined the genetic control of seed dormancy and flowering time in more detail.

4.4.5.1 Potential pleiotropic effects associated with seed dormancy

Based on previous studies, dormancy was expected to be associated with certain testa related traits, potentially including testa thickness (Miao et al., 2001, Murphy and Fuller, 2017, Weeden, 2018), pigmentation (Debeaujon et al., 2000, Weeden, 2007) and testa roughness. In total nine testa-related traits were analysed with 16 independent QTLs tentatively recognised (Supplementary Figure 4.6). This included seven affecting testa structures (five thickness and two roughness loci), seven affecting testa pigmentation and two affecting permeability. All QTLs had the same allelic direction where plants carrying the domesticated allele showed a reduction in dormancy, testa thickness, testa roughness and pigmentation, except for the minor testa mottling *Mt3b*. Our results showed the major permeability, *Perm2* mapped near to major pigmentation and testa thickness QTLs; *A*, and *TT2*, indicating a possible pleiotropic association. A potential relationship among pigmentation, thickness and permeability is supported by induced pigmentation mutants in arabidopsis which show reduced proanthocyanidin expression is positively correlated with reduced testa thickness and increased permeability (Debeaujon et al., 2000, Appelhagen et al., 2014), and similar correlations are also observed in other species (Slattery et al., 1982, Debeaujon et al., 2001, Gu et al., 2011, Bajaj et al., 2015). However, apart from the case *A*/*TT2*/*Perm2*, there were no other instances where additional pigmentation or testa structure QTLs were associated with effects on permeability. This suggests that loci affecting testa thickness do not necessarily affect permeability, a question that will be further explored in chapters 5 and 6.

4.4.5.2 Potential pleiotropic effect of flowering time genes

Flowering time QTLs were found in all four QTL clusters and were considered likely to influence a range of other traits, explaining to some extent the observed QTL clustering. There is growing evidence for the multi-functional role of flowering time genes (Pin and Nilsson, 2012, Tsuji, 2017), including effects on processes such as stomatal opening (Kinoshita et al., 2011), onion bulb formation (Lee et al., 2013), potato tuberisation (Navarro et al., 2011), and stem/inflorescence growth cessation (Böhlenius et al., 2006, Lifschitz et al., 2014). In numerous legume species flowering time has been associated with growth habit and branching (Espinoza et al., 2012, Yang et al., 2017) again suggesting some degree in overlap in the genetic control, either through pleiotropic effect of a single gene or close linkage of numerous different genes. There are traits that will be intrinsically related to flowering time and

therefore expect to share similar genetic control. For example, earlier flowering time enabled the shift from winter to spring cropping by reducing the vegetative phase. As a result, this would likely have repercussion to development rate, plant height and productivity (Dodd et al., 2005, Harmer, 2009, Franklin and Quail, 2010). This is consistent with other crop species where flowering time QTLs were often found clustered with other domestication-related QTLs (Koinange et al., 1996, Isemura et al., 2007, Wills and Burke, 2007, Kaga et al., 2008, Isemura et al., 2012, Kongjaimun et al., 2012, Lo et al., 2018). These associations predominantly involve traits such as plant height, internode length, branching, seed and organ size (Bomblies and Doebley, 2006, Wang et al., 2009, Yu et al., 2015).

To analyse potential pleiotropic effects, a range of traits were scored on a near isogenic line (NIL) population carrying wild *ELF3* allele in a domesticated background. Results showed plants carrying the earlier flowering domesticated allele exhibiting a shorter more robust growth habit with reduced branching and an increased rate of maturity, typical of the *Domestication Syndrome* phenotype. To rule out that these pleiotropic effects were exclusive to the *ELF3* gene, traits were also scored on three Clst6 (*DTF6*) advanced generation segregating populations. Although neither of these populations could definitively rule out the effect of closely linked genes, the nature of the associations was similar in both cases and strengthened the evidence in favour of pleiotropic effects of underlying flowering time genes.

4.5 Conclusions

Crop domestication has had profound effects on plant morphology, but the molecular consequences are poorly understood. This study has provided one of the most comprehensive genetic investigations to date into the domestication of pea. Consistent with previous findings we report domestication related traits are controlled by relatively few but major affecting QTLs and describe an apparent clustering of domestication-related QTLs in the pea genome. This may be due both to pleiotropic effect of flowering time QTLs, and the action of independent closely linked genes, while no obvious clustering was observed around genes controlling what can arguably be considered true domestication traits; Permeability (*Perm2* and *Perm7)* and dehiscent pods (*Dpo1*).

Examining the genetic control of two key domestication traits, two seed dormancy and five flowering time QTLs were identified. For seed dormancy, *Perm2* was considered the critical domestication QTL due to its dominant affect in comparison to *Perm7* which was considered diversification. This was mapped near a pigmentation and testa thickness QTLs which indicates these traits may be linked, this will be explored in chapter 5 and 6. For flowering time, three novel QTLs were identified from previous examination of this cross in SD conditions, with one of these (*DTF3b*) likely representing the effect of

the to the induced *le-3* mutation provided from the NGB5839 parent. Chapter 7 will explore in detail the genetic control of *DTF2* and *DTF5*.

This study was based on a single biparental cross between a wild accession and a highly adapted, modern cultivar. As such it would be expected to detect not only domestication traits, but many other traits arising more recently through subsequent diversification and breeding. To clarify these distinctions, future research should undertake a similar analysis in wild x landrace and landrace x modern cultivar populations. In addition, examining genetic control in diverse populations such as *P. sativum var abyssinicum* would explore different evolutionary trajectories and may identify rare beneficial alleles of known genes, which could be utilised in plant breeding programs.

Chapter 5: Investigating seed dormancy related QTLs on LGII

5.1 Introduction

Seed dormancy is the phenomenon in which seeds are maintained in a reversible state of reduced metabolic activity. It is effectively a bet-hedging strategy to prevent or delay germination in sub-optimal growing conditions (Childs et al., 2010) that contributes to the population's long-term survival. Dormancy allowing seeds to remain viable within the seedbank and either germinate synchronously when conditions are more favourable, or germinate only sporadically, thus spreading the risk against catastrophic loss that can potentially result from the fluctuating conditions commonly encountered in nature.

In peas and many other legume species, dormancy is primarily maintained by the testa, which acts as an impermeable physical barrier to water uptake (Werker et al., 1979, Smykal et al., 2014, Hradilová et al., 2017). This impermeability is thought to be determined by the structural and chemical composition of the testa, which is described in detail in Chapter 1. Comparison of the two parent lines used in this study, provides an illustration of the overall effects of domestication, which include the loss of conspicuous testa roughness, a decrease in testa thickness, and a reduction in pigmentation (Chapter 4), findings which are consistent with other reports (Zohary and Hopf, 1973, Werker et al., 1979, Weeden, 2007, Smykal et al., 2014, Hradilová et al., 2017). Despite the existence of these distinct morphological differences between wild and domesticated lines, little is known about their genetic or molecular basis (Smykal et al., 2014). As part of the QTL analysis performed in chapter 4, the molecular control of permeability and other testa traits potentially related to seed dormancy were analysed (Figure 5.1). Two significant permeability QTLs; a major QTL on LGII (*Perm2*) and a minor QTL on LGVII (*Perm7*). *Perm7* mapped to a novel region and was not associated with either testa thickness or pigmentation, while *Perm2* was associated with loci for two additional testa related traits, the testa thickness QTL *TT2* and Mendel *A* locus, which influences pigmentation of the seed coat, flowers and stems (Hellens et al., 2010).

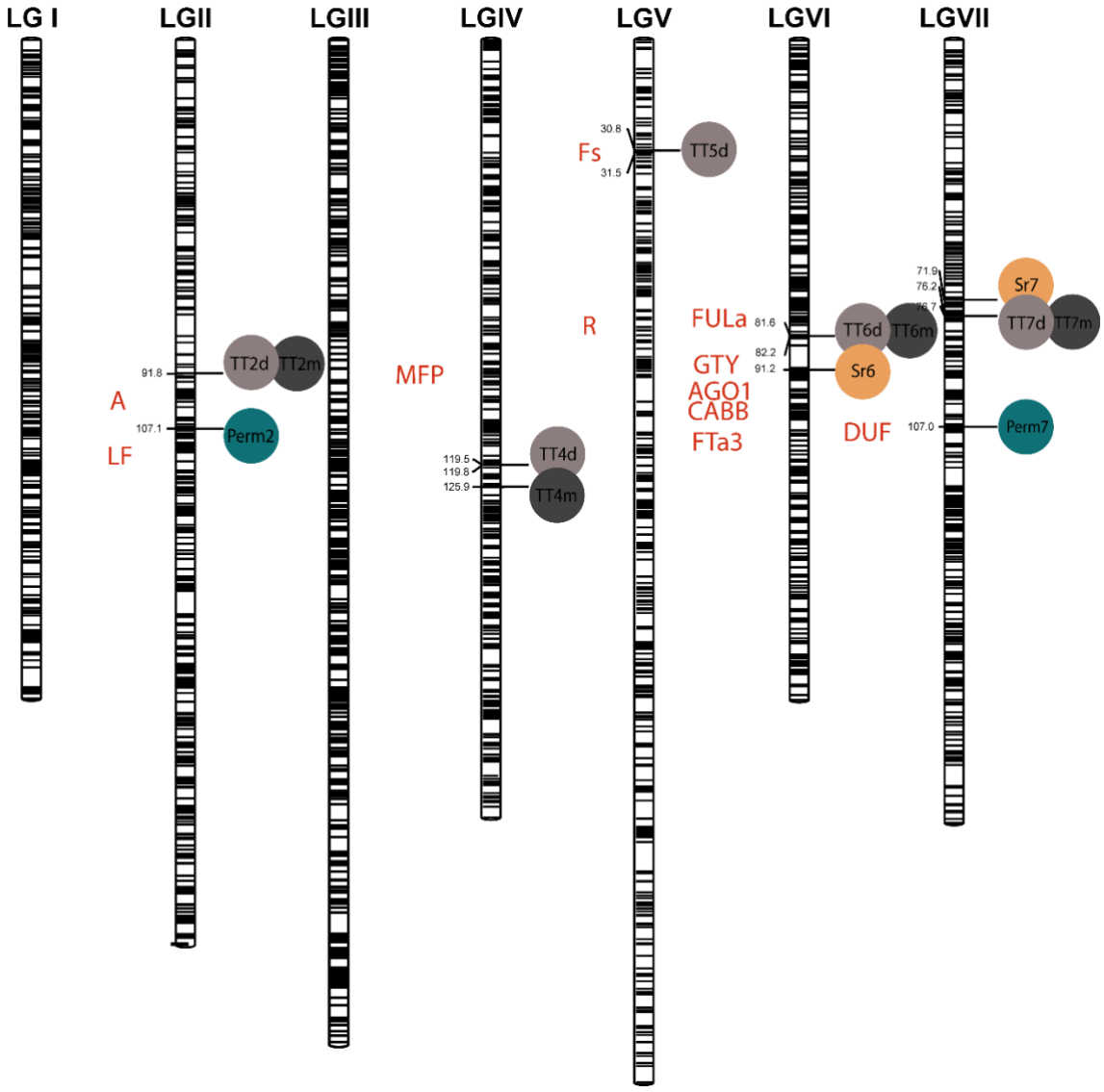


Figure 5. 1: Schematic representation of seed dormancy (Perm) and other potentially related traits testa thickness (TT) and seed coat roughness (Sr). Anchor markers and known Mendelian loci are indicated on the left of each linkage group.

Along with indehiscent pods, loss of seed dormancy has been recognised as fundamental to the domestication of pea and other legumes crops (Ladizinsky, 1987, Abbo et al., 2011). It was therefore expected that any major variants selected during domestication should be represented in the JI1794 x NGB5839 population, as well as other subsequent modifications. As discussed in Chapter 4, *Perm2* was tentatively considered a critical domestication QTL due to its dominant effect on seed dormancy and similar reports in other populations (Weeden, 2007). In comparison *Perm7* had a less significant effect, and to our knowledge previously not been reported, which collectively suggests this is a diversification QTL. *Perm2*, was therefore of interest to this study due to it potentially specifying an important early reduction in physical dormancy. Furthermore, its colocation with loci controlling pigmentation (*A*) (Hellens et al., 2010) and tentative mapping of the testa thickness locus *EP*

(Kaznowski, 1926, Wade, 1937, Blixt, 1972) may be indicative of pleiotropic effect, however there is no molecular evidence that links all three QTLs together.

The concurrent loss of pigmentation and thinning of the testa can be observed in numerous crop species including *L. culinaris* (Matus et al., 1993), *C. arietinum* (Penmetsa et al., 2016), *Brassica napus (Qi et al., 2017)* and *Brassica rapa* (Li et al., 2012) and would seem to implicate Mendel *A* gene (a bHLH regulator of anthocyanin biosynthesis) as a potential candidate for *Perm2*. This is further supported by the fact that mutants for the orthologous gene in *arabidopsis* reduced testa thickness and increased permeability (Debeaujon et al., 2000, Appelhagen et al., 2014). In contrast, pigmentation loss in the domesticated rice bean (Isemura et al., 2010), Mungbean (Isemura et al., 2012) and *Trifolium subterraneum L*. (Slattery et al., 1982) showed no significant effect on permeability. Furthermore, the presence of pigmented seedcoats in numerous pea landraces and other domesticated material (Holdsworth et al., 2017) strongly argues against a central role for the loss of *A* function with the domestication-related loss of dormancy, and implies some other alternative dormancy-regulating mechanism (Smykal et al., 2014, Hradilová et al., 2017).

The first aim of this is to validate seed dormancy *Perm2* and testa thickness *TT2* originally identified in chapter 4, using advanced generation segregating populations. Secondly, we intend to examine the genetic relationship among testa traits potentially related to seed dormancy and domestication.

5.2 Methods

5.2.1 Plant material

5.2.1.1 Developing advanced generation segregating populations

Advanced generation segregating populations were developed from the RILs F_2 progenitor population. These were selected on the basis that the Cluster 2 region (see chapter 4) was segregating, but as far as possible, given the available genotype combinations in the F_2, testa thickness QTLs outside the Clst2 region were fixed – a method similar to (Peleman et al., 2005). Identification of markers from the F_2 population to represent peak QTL markers was achieved by comparing their relative positions in *M. truncatula* genome using the genomic resource platform Phytozome v.12 (Goodstein et al., 2012). When possible two or more markers were used to flank QTL region, to better account for uncertainty in the precise QTL position.

5.2.1.2 NILs for the a2 mutant

The *A2* gene encodes a WD40 repeat protein (Hellens et al., 2010) which, like Mendel *A*, is a component of the MYB-bHLH-WD40 (MBW) transcription factor complex that has a fundamental and

highly conserved role in regulation of the flavonoid biosynthesis pathway in plants (Xu et al., 2015, Li et al., 2016). The wild-type (A2) line JI2822, and two independent induced *a2* mutant lines JI3559 (*a2-3*) and JI3560 (*a2-4*) (Hellens et al., 2010) were obtained from the collection at the John Innes Centre in Norwich, UK.

5.2.2 Growth conditions

All populations were grown at the University of Hobart using protocols described in Chapter 2.

5.2.3 Phenotyping

Testa thickness was measured using micrometer, as analyses in chapter 4 indicated that this method resulted in less error than the use of digital calipers, and could therefore considered the more accurate and reliable technique. Measuring techniques for testa thickness and permeability followed the same protocols described in chapter 4.

5.2.4 Genotyping

5.2.4.1 Development of markers

Nine new and existing markers (Table 5.1) were used to increase resolution within the Clst2 region, spanning 38.4cM around the *A* gene according to the map presented in Chapter 3 (PsCam009518 73.640cM to LF 122.000cM). *LF* and Mendel *A* are pre-existing markers used in the development of the RILs map. All remaining markers were developed from pea transcript sequences with selection guided by DarT marker positions in RIL map (Chapter 3) or the high density consensus map developed by Tayeh et al. (2015).

*Table 5. 1: Details of target genes and new HRM markers developed for the analysis of the Clst2 region, *closest DarT marker and approximate position, ** closest PsCam to DarT marker mapped in Tayeh et al. (2015), *** position on map unknown.*

Marker	Type of marker	Forward and reverse sequence	Tm (°C)	Mapped position in *M. truncatula*	Mapped position in Pea (Tayeh et al., 2015)	DarT no.	RILs Map position
PsCam009518	Size	F: TTGGACTGGTTGATGAGTGG	58.4	1g046620	27	3563381	73.640
		R: TCCATCAATCCATTCTGTCG	56.4				
Cwf	HRM	F: TTGTCAGTATCCTAATAGTTTGA	55.7	1g077570	31.4	3555339	89.262
		R: CCAATCCATTGTTATGTCTCC	57.4				
PsCam056891	Size	F: AATCCACCTGTGGAACTCG	57.3	1g111970	33.3	***	***
		R: TTTGAGAGTGGCTAAACATGG	57.4				
Mtran	HRM	F: TGATCAGTCGCCTCATCAGC	60.5	1g071110	34.2	3564794*	96.500*
		R: AATGTTGGGTGGACAGGACC	62.5				
NEK4	HRM	F: GAAATGGCTGCGCACAAACC	60.5	1g071480	34.4	3537228*	96.848*
		R: CCTAGCATGGTTCGAGTCGG	60.5				
Mendel *A*	HRM + Phenotype	F: TCCAATCGAAGAACCTCTCG	58.4	1g072320	35.6	A	99.627
		R: GGGTTAGGAGTTAGGACAAACC	62.1				
ArgJ	HRM	F: AAAAATCCGAGGGCAAGATA	54.3	1g068845 (1g068825)**	38.4**	3562317	104.035
		R: CTGTGGATAGATGAGACTTGCAT	61.1				
NAD	HRM	F: GGGGTTTGCTGAACACATTA	56.4	1g067530	38.8	3563452	106.578
		R: TGGGTGCACAAGAGGAATAA	56.4				
LF	HRM	F: GGTCCCTCTTTACCCTGGTATT	62.1	1g060190	44.3	LF	112.000
		R: TGATCTGCAGGAAAACAATAAA	54.7				

5.3 Results

5.3.1 Phenotypic analysis of RILs

As previously discussed in chapter 4, a major seed permeability QTL was identified on LG II. As shown in Figure 5.2 a detailed inspection of this region shows this mapped near two anchor markers *LF* and *A*, and included two additional seed coat related traits, testa thickness (*TT*) and pigmentation (*A*).

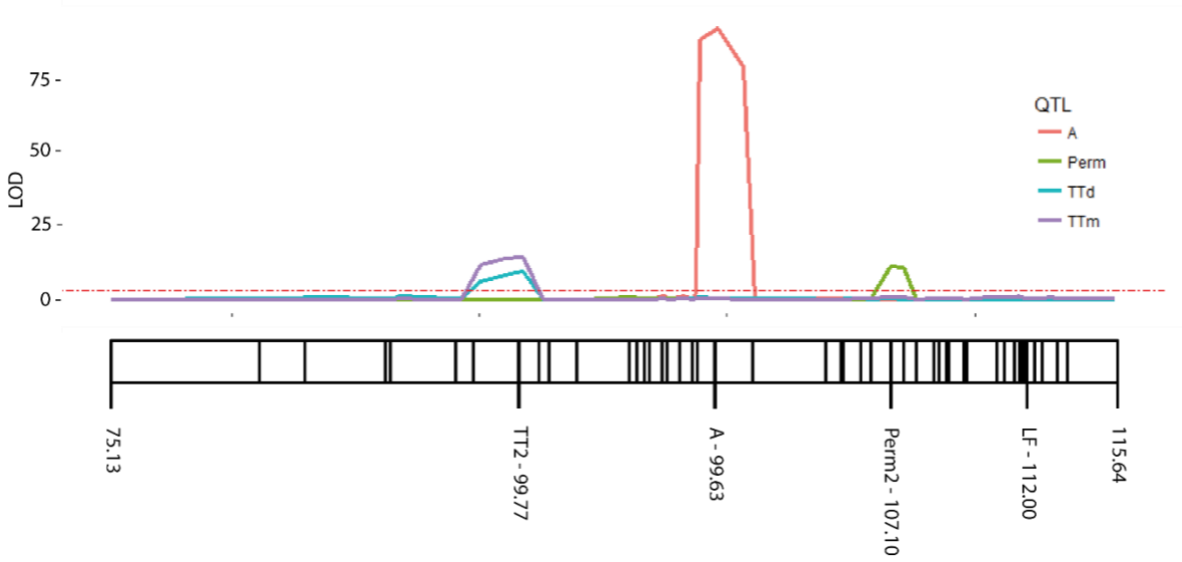


Figure 5. 2 Detailed inspection of Perm2 region which includes QTL analysis data collected from chapter 4 testa thickness using digital caliper (TTd) and micrometer (TTm) and seed coat permeability (A). Region was analysed between 75.13cM to 115.64cM on PsLGII, markers positions across region are shown by black lines, with anchor and peak markers positions given. Dashed red line shows the LOD3 significance threshold.

Using pigmentation to indicate genotype at Mendel *A* locus, a strong correlation with both testa thickness and permeability were observed (Figure 5.3a). Results showed the domesticated genotype (represented by JoinMap code as "B") generally had a thinner and more permeable testa than the wild (JoinMap code "A"). This was also apparent in the pairwise analysis shown in Figures 5.3b and 5.3c. This confirms that the *A* region did had a significant effect on both testa thickness and permeability when measured across the RIL populations, regardless of influences from additional genomic regions.

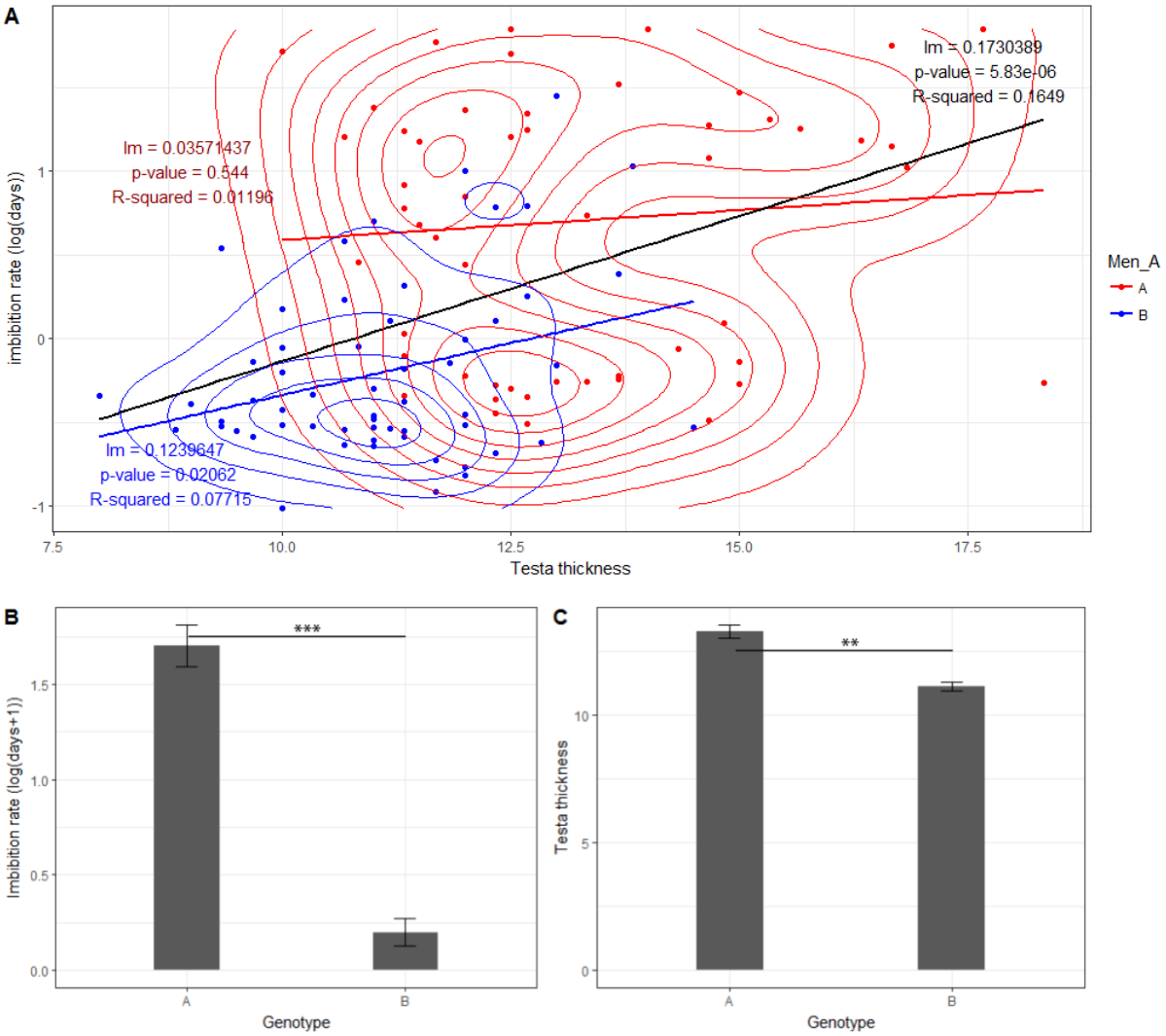


*Figure 5. 3: Comparing testa thickness and permeability in wild and domesticated Clst2 genotypes across RILs population A) Density plot between testa and permeability, B) Bar graph of permeability, C) Bar graph of testa thickness. Significant difference between wild and domesticated alleles conducted using Tukeys test (NS >0.05, * P<0.1, ** P<0.01, *** P<0.001)*

This is represented in another way in Figure 54, which indicates the distribution of trait scores for each individual RIL. Again, it is clear that the Clst2 region showed a strong association with both permeability and testa thickness, but the distribution of scores for lines carrying domesticated and wild alleles was not consistent with control by a single segregating locus but showed substantial overlap through the middle of the distribution. This indicates that additional loci affecting these traits are segregating within the population, in keeping with the QTL analysis in chapter 4.

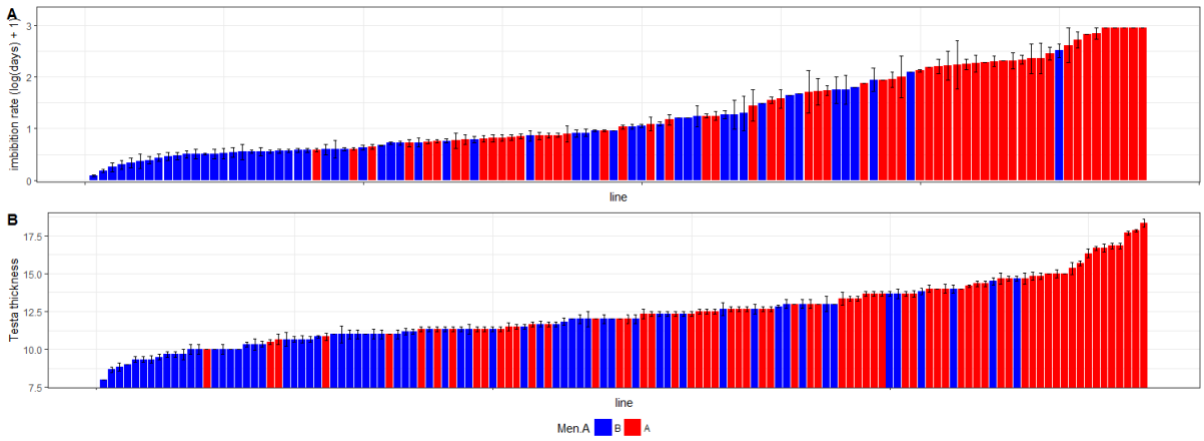


Figure 5. 4: Bar graph for A) permeability and B) testa thickness across RILs population. Lines are ordered according to value.

5.3.2 Developing advanced generation segregating populations

In an attempt to analyse the Clst2 region as a Mendelian locus for testa thickness and permeability, several different segregating populations were identified from the F_2 generation of the derivative RIL population, which had been genotyped for a range of gene-based markers to a varying degree of coverage across the genome. The aim was to select lines that segregated for the Clst2 region but were fixed for additional testa thickness and permeability QTLs wherever possible.

Representative markers for each QTL were identified by comparing the relative position of QTL peak markers and other gene-based markers in the *M. truncatula* genome. As shown in Table 5.2a and visualised in Figure 5.1, representative markers were identified for the five testa thickness QTLs; *TT2*, *TT5*, *TT6* and *TT7*. The two markers *A* and *LF* spanned the peak marker for *TT2*. Two markers were identified for *TT6*, which although unable to flank its peak marker did provide conservative size QTL. A single representative marker was used for *TT7* which mapped near peak marker and a phenotypic marker conferring to blackspots (*Fs*) was identified for *TT5*, by phenotyping in the F_3 population. No representative markers were available for *TT4*. Representative markers were identified for *Perm2* and *Perm7* (Table 5.2b), with the same two markers *A* and *LF* used to indicate the presence of *Perm2*. Only a single marker mapping close to the *Perm7* peak marker was available. As shown in Table 5.3, around 40 to 46% of the observed ~50% testa thickness observed phenotypic variation was accounted for, depending on genotyping of *TT5*. Whereas for permeability the entire 41.3% observed phenotypic variation could be accounted for, although the possibility of potential undetected recombination between *Perm7* and its representative marker, *DUF* could not be excluded.

As shown in Figure 5.5, two F_3 lines were grown initially; TT2-3A which had a fixed wild background (n=17) and TT2-3B which was fixed for the domesticated background (n=7). Because of their small size

these populations were genotyped for the representative markers *A* and *LF* and six heterozygous individuals were advanced to the F_4 to generate larger segregating progenies.

Table 5. 2 Table detailing representative markers identified in the F_2 population for genotyping a) Testa thickness and b) Permeability QTLs for the development of the advanced generation segregating population. Peak QTL markers and F_2 representative markers positions were determined by relative position in the M. truncatula Mt4.0 V1 reference genome. Marker positions are represented as gene number and the start Bp position. If the QTL peak marker did not map in Medicago, next closet marker was used.

a)

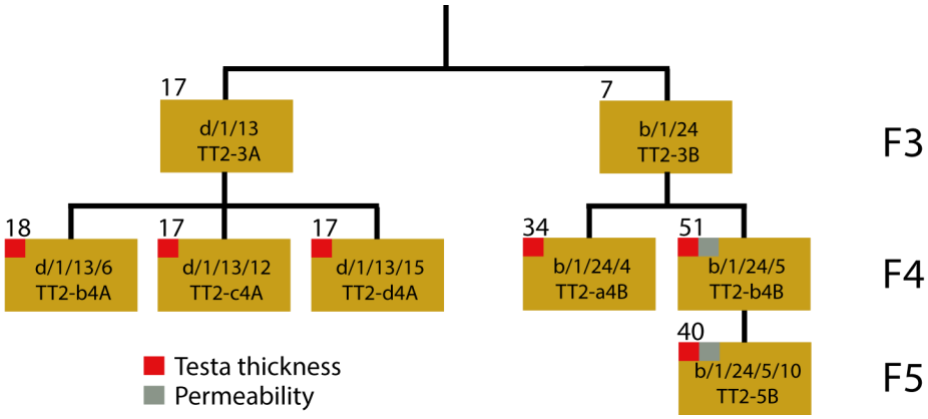
	TT2		TT4			TT5		TT6			TT7		
QTL	Peak DarT marker	Medicago Position	Peak DarT marker		Medicago Position	Peak DarT marker	Medicago Position	Peak DarT marker		Medicago Position	Peak DarT marker		Medicago Position
	3556039_1	1g070505 chr1 31.30Mb	TT4d	4663134_3	4g099070 chr4 41.02Mb	3560771_3	7g113410 chr7:46.71Mb	TT6d	3543436_3	6g075160 chr6:27.83Mb	TT7d	3542996_3	4g094625 chr4:38401265
			TT4m	4655522_1	4g101930 chr4 42.17Mb			TT6m	3561639_2	6g072705 chr6:26.93Mb	TT7m	4662724_3	4g094305 chr4:37676691
Representative marker	Marker	Position	none			none		Marker		Position	Marker		Position
	LF	1g60190 chr1 26.20Mb						MLO1		6g033330 chr6 10.66Mb	SVPb		4g093970 chr4:37162662
	Mendel A	1g072320 chr1 32.09Mb						RNAhel		6g056080 chr6 20.12Mb			

b)

	Perm2		Perm7	
QTL	Peak DarT marker	Medicago Position	Peak DarT marker	Medicago Position
	5251908_1	1g067530 chr1 29.13Mb	3542137_3	4g083460 chr4 32.47Mb
Representative marker	Marker	Position	Marker	Position
	LF	1g60190 chr1 26.20Mb	DUF	4g083440 chr4 32.47Mb
	Mendel A	1g072320 chr1 32.10Mb		

Table 5. 3: Explained phenotypic variation for represented and non-represented testa thickness QTLs used in the development of the advanced generation segregating populations. Values based from results in chapter 4

	PEV%				
	Testa thickness			Permeability	
	QTL	Testa thickness Digital Caliper	Testa thickness Micrometer	QTL	Permeability
Represented	TT2	16.6	29.1	Perm2	32.7
	TT6	12.3	5.5	Perm7	8.6
	TT7	11.5	5.7		
	PEV%	***40.4***	***40.3***	***PEV%***	***41.3***
Partly represented	TT5	6.3	NA		
	PEV%	***6.3%***	***NA***		
Not represented	TT4	6.7	9.4		
	PEV%	***13***	***9.4***	***PEV%***	***0***
Total		***53.4***	***49.7***		***41.3***



			QTLs and representative markers							
			TT2/Perm2		TT5	TT6		TT7	Perm7	
Parental pop	Population	Generation	LF	A	Fs	RNAhel	MLO	SVPb	DUF	Number of seeds
d/1/13	TT2-3A	F_3	H	H	H	A	A	A	A	17
d/1/13/6	TT2-b4A	F_4	A	H	H	A	A	A	A	18
d/1/13/12	TT2-c4A	F_4	A	H	H	A	A	A	A	17
d/1/13/15	TT2-d4A	F_4	H	H	H	A	A	A	A	17
b/1/24	TT2-3B	F_3	H	H	A	B	B	B	B	7
b/1/24	TT2-a4B	F_4	H	H	A	B	B	B	B	34
b/1/24	TT2-b4B	F_4	H	H	A	B	B	B	B	51
b/1/24	TT2-5B	F_5	H	H	A	B	B	B	B	40

Figure 5. 5: schematic diagram illustrating the development of the TT2 advanced generation segregating populations from the F2 to F5 populations. Populations measured for testa thickness and/or permeability are indicated with a red and grey box respectively

5.3.3 Improving resolution around *TT2*

The F_4 populations were phenotyped for testa thickness and pigmentation and genotyped for *LF* (Figure 5.5). Figure 5.6 shows that plants homozygous for the domesticated allele of *A* or *LF* had significantly thinner testas than those carrying at least one wild allele. Exceptions were seen with population TT2- b4A for markers *A* and *LF* and TT2-c4A for marker *LF* only. Collectively this showed a significant difference in testa thickness between genotypes, thereby validating *TT2*.

The Tukeys pairwise analysis showed the Mendel *A* marker had a more significant difference between genotypes than the *LF* marker, at least in progenies TT2-c4A and TT2-d4A suggesting that it might be more informative and potentially closer to the causal gene. This was further supported by the fact that progeny TT2-b4A which was segregating for *LF* but not for *A,* also did not appear to show any difference in testa thickness associated with *LF*. This progeny therefore appears to be carrying a recombination that separates the *TT2* effect from *LF* and that would provide an upper boundary to the *TT2* position.

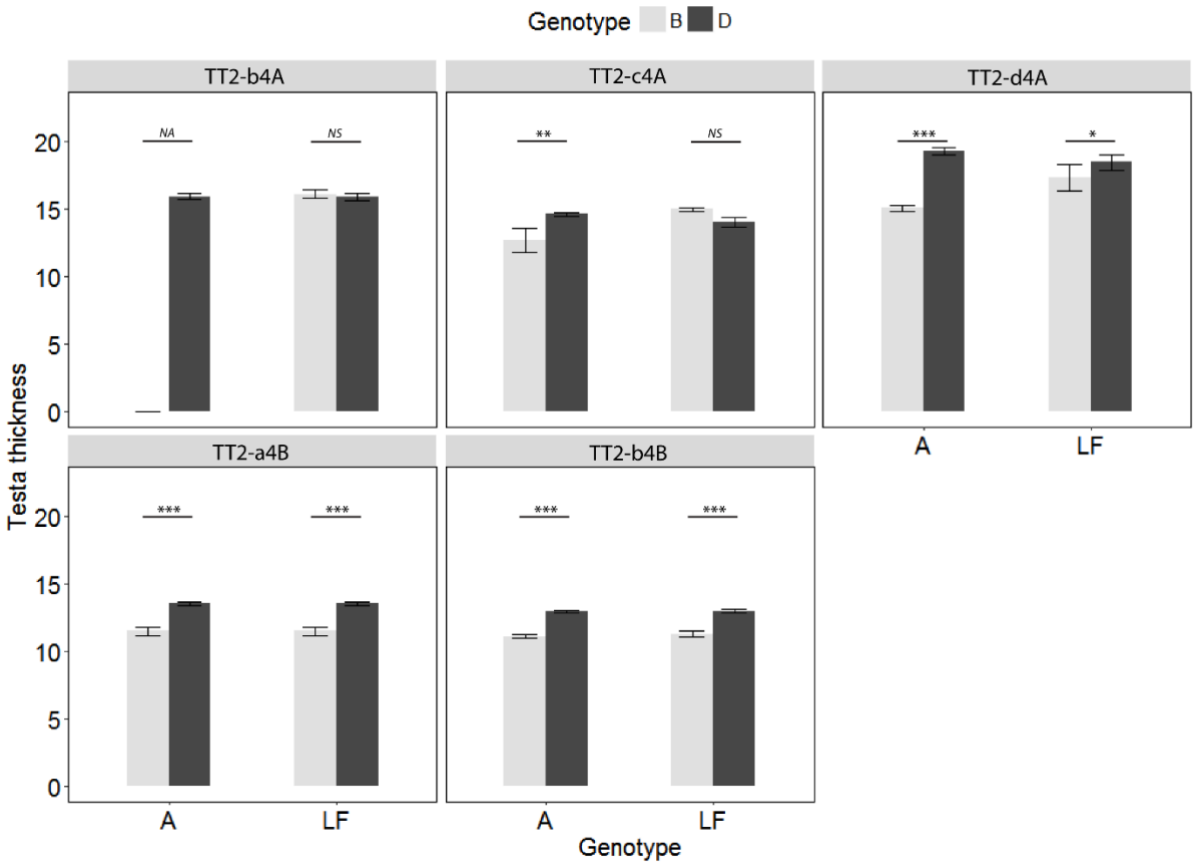


*Figure 5. 6: Comparing testa thickness across F4 advanced generation segregating populations using Clst2 markers: Mendel A and LF. Populations TT2b-4A, TT2c-4A, TT2d-4A are fixed with wild background and TT2a-4B and TT2b-4B are fixed with domesticated background. Significance analysed using Tukeys pairwise analysis (NS >0.05, * P < 0.05, ** P<0.01, ***P<0.001)*

To better define the recombinations in this region, several additional markers were then developed, including *A* (genotype), *ArgJ*, *Cwf18*, *Mtran*, *NAD*, *NEK4*, *PsCam009518* and *PsCam056891* (Table 5.1). These markers were mapped in the largest progeny, TT2-b4B (n=51), and in TT2-b4A which, as described above, appeared to be fixed for *TT2* and for *A*.

Markers were ordered according to recombination events, as shown in Figure 5.7. Pairs of markers showing no recombination were ordered based on data from the RIL population or the linkage map developed by Tayeh et al. (2015). Results showed marker order was consistent with the *M. truncatula*

reference genome and where applicable with Tayeh et al. (2015) pea linkage map, supporting the overall robustness of the analysis

As shown in Figure 5.7a population TT2-b4B was found to be fixed with wild genotypes for all markers below *A* (i.e. between *NEK4* and *PsCAM009518)* but confirmed to be segregating for both *A* and *LF* markers. This observation defines NEK4 as the lower QTL boundary.

Figure 5.7b shows that *A* was still segregating in the TT2-b4A progeny despite previous assumptions this had become fixed due to all lines exhibiting pigmented phenotype. Results showed this incorrect conclusion resulted from the scoring of A as a dominant classical marker in the previous (F_3) generation and the apparent absence of domesticated alleles between *NAD* and *ArgJ*. Based on collective findings in Figure 5.6 which indicated population was not segregating for testa thickness and tended to show the domesticated allele is recessive, this tentatively suggests *NAD* as the upper QTL boundary. Note, this is based from a limited population size and does not rule out the possibility that the causative gene is situated outside the *LF* in the *LF-NEK4* region.

Despite not providing a conclusive upper limit on the location of *TT2*, the two progenies together identified potential novel recombinations that should allow an upper limit to be determined in future. Lines TT2-b4A-4/11 and TT2-b4B-5 both appeared to have sustained recombinations between *NAD* and *LF*, while lines TT2-b4B-47 and TT2-b4A-15 could potentially be used to narrow the lower boundary and definitively exclude A as a candidate. Caution must be taken with the genotyping of *A* in line TT2-b4A-15, which suggests a double recombination event between *NAD* and *A*.

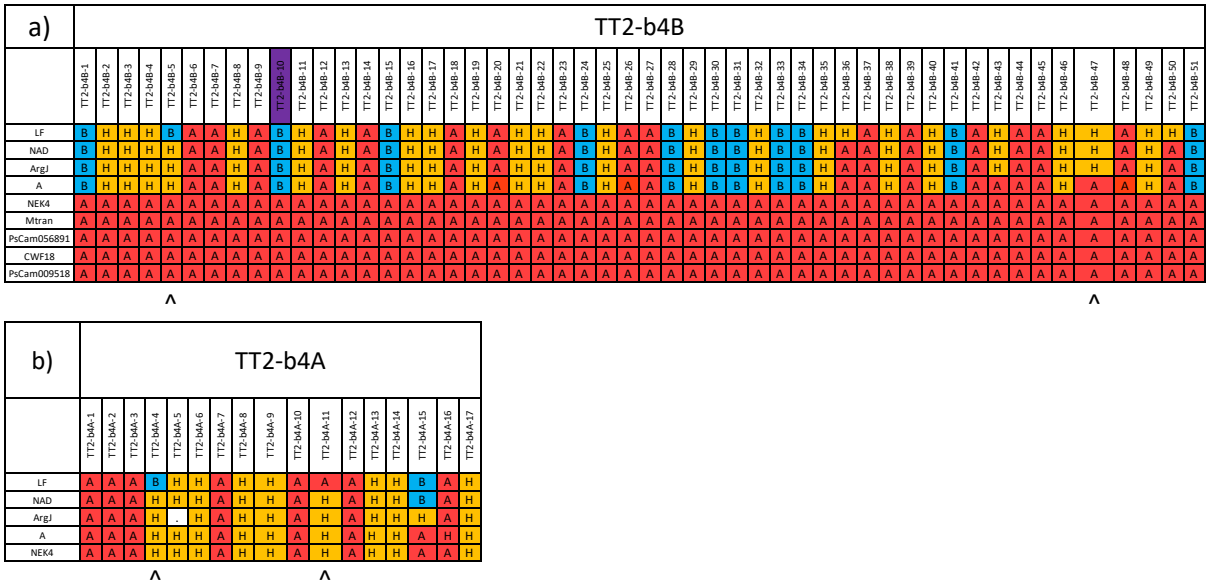


Figure 5. 7: Genotyping of new markers across populations TT2-b4A and TT2-b4B population. Marker order was based on best fit. Where order could not be established due to lack of recombinations markers were ordered based on RIL or previous linkage maps. Colours represent different genotypes red = Wild, blue = domesticated and orange = heterozygous. Line highlighted in purple represents TT2-b4B-10 which was advanced to an F_5 population. Potential lines that could be advanced to TT2 and Perm2 interval highlighted by the ^.

5.3.4 Fine mapping *Perm2*

In a similar attempt to validate *Perm2*, the TT2-b4B progeny was also assessed for permeability. As previously discussed this population was fixed for other testa thickness and permeability QTLs. Figure 5.8 shows that plants homozygous for the domesticated allele had testas that were on average statistically more permeable than homozygous wild or heterozygous lines, confirming an effect of this region on permeability. No significant difference was observed between wild and heterozygous, which indicates that the increased permeability conferred by the domesticated allele was inherited in a fully recessive manner.

As in the case of testa thickness discussed above, *NEK4* could also be defined as the lower boundary for *Perm2*. On average, plants carrying the domesticated Mendel *A* marker were most permeable, and this effect decreased incrementally across adjacent markers (Figure 5.8). Although this might be taken to imply that Mendel *A* is the most informative marker, the difference is not statistically significant, and when actual recombinations within the region are considered (Figure 5.7), it is clear that no recombinations distinguish *A* from its adjacent marker ArgJ in the F_4. However, as for *TT2*, recombinants in the TT2-b4B progeny may allow an upper boundary for *Perm2* to be determined in the F_5 generation.

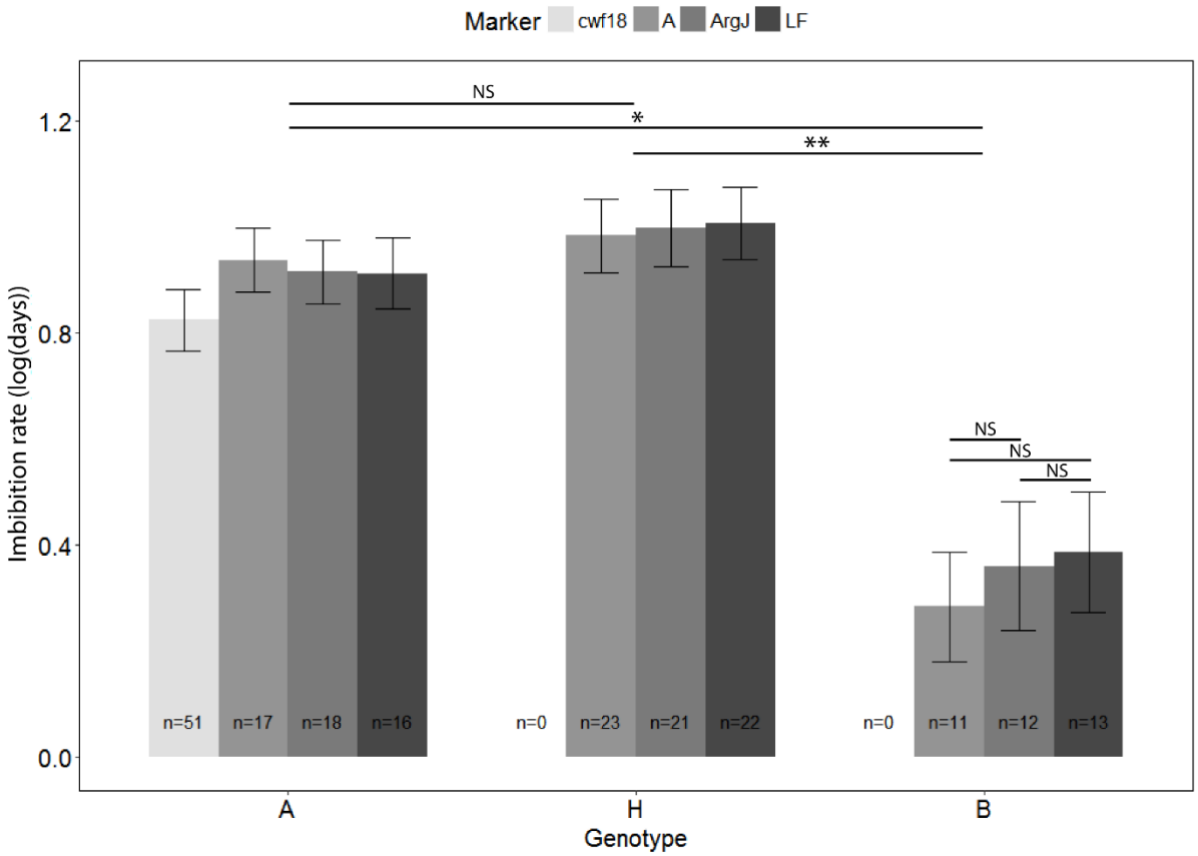


*Figure 5. 8: Fine mapping of permeability in domesticated background TT2b-4B. A) Bar chart showing average permeability for each genotype across all markers. Due to identical genotyping between markers PsCam to NEK4 and ArgJ to NAD (Figure 5.7) these were referred to as Cwf18 and ArgJ respectively. Using Mendel A results significant difference between genotypes was calculated using Tukeys pairwise analysis (NS >0.05, * P < 0.05, ** P<0.01, ***P<0.001). No significant difference was observed between markers of same genotype, data not shown.*

5.3.5 Pigmentation as candidate for *TT2* and *Perm2*

As already established by the analyses above, Figure 5.9 also emphasizes the strong correlation of the genotype at the *A* locus with testa thickness and permeability. A similar analysis was also conducted with *ArgJ* and *LF*, but the correlation was less strong (data not shown). This again implies that *A* could be a strong candidate gene.

However, the data for line TT2-b4B-10 provide evidence against this, since this line was homozygous for the domesticated (*a*) allele but exhibited a relatively impermeable phenotype characteristic of the homozygous wild or heterozygous genotype (Figure 5.9). Because this line was fixed between *LF* and *A* for domesticated allele and fixed between markers *NEK4* and *PsCam009518* for wild allele (see Figure 5.7) this suggests causative gene might be located between *A and NEK4*, implying an order *NEK4-Perm2-A.* However, this would require an additional recombination to have occurred in this individual between *Perm2* and *A*, and would predict *Perm2* to be segregating in the next generation. This interpretation was explored by advancing line TT2-b4B-10 to the F_5 generation, and phenotyping population for permeability. As shown in Figure 5.9, permeability was not segregating in this

population, but exhibited a phenotype reminiscent of the domesticated genotype. The apparently recombinant, low-permeability phenotype of TT2b-4B-10 was thus considered an anomalous result, without obvious explanation. Nevertheless, the F_5 result reinstates the *A* gene as a possible candidate for both testa thickness and permeability that should be explored further in future by generation of larger populations and identification of additional relevant recombinations.

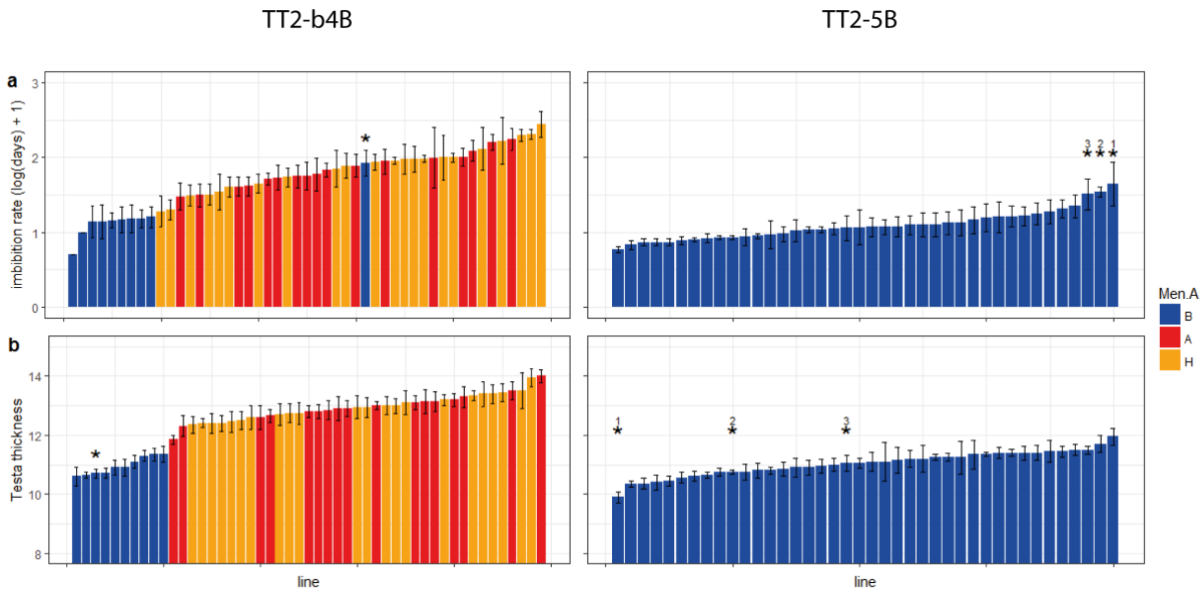


*Figure 5. 9: Analysing a) permeability and b) testa thickness in relation to Mendel A marker across populations TT2-b4B and population TT2-5B which was advanced from TT2-b4B-10, highlighted by the *. Corresponding testa thickness of the top 3 most impermeable lines in population TT2-5B are denoted with the *(1-3) symbol*

5.3.5.1 Reverse genetics approach

As the *a* mutation and the putative causal mutations for *TT2* and *Perm2* have arisen naturally, it is difficult to separate them genetically. One way to further explore this question would be to examine the effects of an induced loss-of-function mutant for *A*. Unfortunately, no such mutant has been reported, however two induced mutants in the *A2* gene have been described, that, like the *a* mutation, eliminate all pigmentation in flowers and seeds (Hellens et al., 2010). *A2* encodes a WD40 protein that acts with the *bHLH* protein encoded by *A* in the so-called "MBW" complex to regulate the expression of genes in the flavonoid pathway (Baudry et al., 2004, Hellens et al., 2010, Li et al., 2016). Although different components of the MBW complex have been shown to regulate different parts of the flavonoid biosynthesis pathway in arabidopsis, similar effects on permeability and testa thickness were observed (Debeaujon et al., 2000, Appelhagen et al., 2014).

Figure 5.10 shows that testas of seeds from two a2 mutant lines (a2-3 and a2-4) were on average significantly thinner than their isogenic wild type. This could imply that some

compound(s) in pathways controlled by the MBW regulatory complex might act to influence testa thickness, and therefore implies that an a mutation might also have a similar effect. However, such a conclusion assumes that the A and A2 genes have no individual roles independent from the MBW complex, but this possibility cannot yet be excluded. In contrast, neither of the induced *a2* mutants showed a significant difference in permeability compared to the wild type. Although this indicates that *a2* does not affect permeability, it is notable that both wild and induced mutant lines exhibited a highly permeable phenotype. This might reflect the fact that both are carrying the domesticated allele of *Perm2*, or alternatively, that effects of other loci elsewhere in the genome on permeability might mask any potential effect of the *A2* gene or make the parent line JI2822 relatively permeable despite carrying the wild *Perm2* allele.

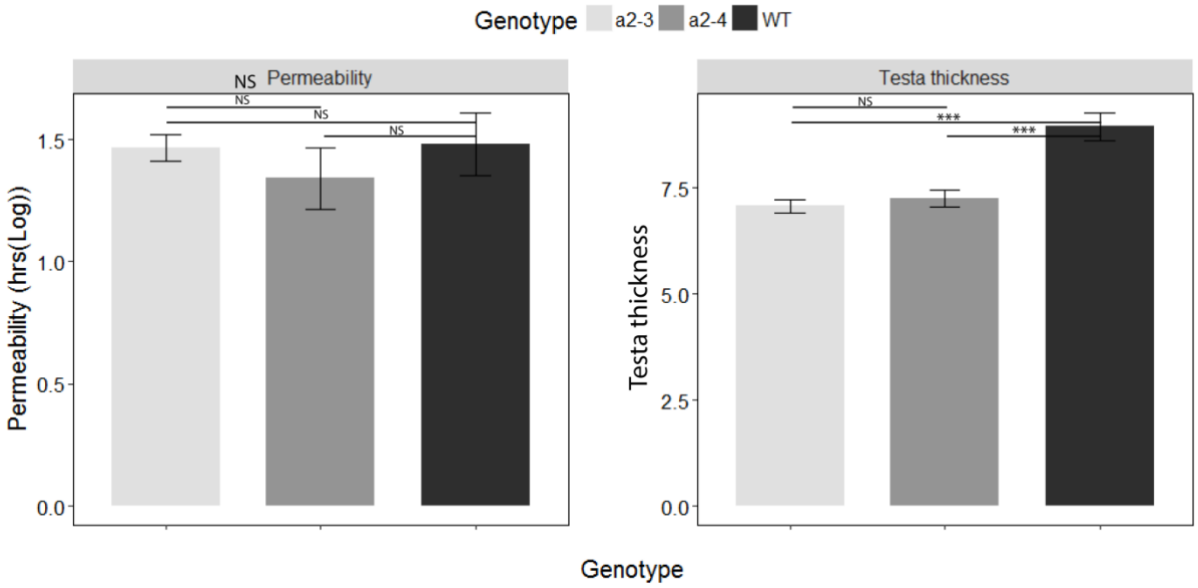


*Figure 5. 10: Comparison between testa thickness and permeability between induced pigmented WT and non-pigmented induced A2 mutants a2-3 and a2-4. Significant difference between lines were calculated using Tukeys pairwise analysis (NS >0.05, * P < 0.05, ** P<0.01, ***P<0.001)*

5.4 Discussion

In chapter 4, major permeability and testa thickness QTL were mapped on LGII near Mendel *A* gene, which encodes a *bHLH* protein essential for anthocyanin accumulation in the seed coat. Because of its dominant effect to permeability and the apparent concurrent reduction with testa thickness, it is possible that this locus might represent the critical mutation leading to the reduction in germination that occurred during domestication of *P. sativum var sativum* lineage. This is consistent with previous findings which also identified testa thickness and permeability QTLs in this region in *wild x cultivated* (MN313 × JI1794) and potentially *primitive × wild* (WL808 × JI261) and *cultivated × primitive* (Cultivated × WL808), which mapped to the testa thickness *EP2* locus on LGII but its position in this

group could not be resolved (Weeden, 2007). However, the causative gene or genes for these traits have not been identified.

To ensure the observed testa thickness and permeability QTLs in this region were real, advanced generation segregating populations were developed. Both QTLs were validated, with plants carrying the domesticated genotype for this region exhibiting significantly thinner and more permeable testa. Fine mapping of testa thickness and permeability QTLs on population TT2-b4B identified *NEK4* as a lower boundary for both loci. This was based from an ancestral recombination event between *A* and *NEK4* that fixed the progeny with the wild genotype between markers *NEK4* and *PsCam009518* while maintaining heterozygosity for markers *A* and above (including *TT2* and *Perm2*). From population TT2-b4A, *NAD* was recognised as a potential upper QTL boundary for testa thickness but was not examined for permeability. This was shown by testa thickness not segregating across populations which is likely due to the absence of lines carrying the recessive domesticated allele between markers *ArgJ* and *NEK4*. Because of its limited population size and an inability to show the QTL was not outside *LF* boundary, the validity of *NAD* as the upper QTL boundary was circumspect, therefore further analysis is required.

The interval between *NEK4* and *NAD* was calculated at 4.4cM based on linkage map developed by Tayeh et al. (2015) consisting of around 62 around genes based on synteny with *M. truncatula*. As mentioned above, *A* remains a strong candidate for both QTLs, with a loss of function mutant allele known to be segregating within the population. However, the culmination of the domestication/diversification process has led to the loss of pigmentation in many crop species. This has been linked to numerous agronomically important traits such as loss of seed dormancy and longevity (Werker et al., 1979, Harris, 1987), as well as recent findings showing loss of pigmentation increased seed size, oil and protein content (Chen et al., 2014, Qi et al., 2017, Lei et al., 2018). Pigmentation is synthesised from diverging branches within the flavonoid biosynthesis pathway which includes flavanols, anthocyanins and proanthocyanidins (PA) (Saito et al., 2013), see chapter 1 for details flavonoid biosynthesis regulatory pathway. In arabidopsis, pigmentation has been directly linked to testa thickness and permeability using induced flavonoid biosynthesis mutants (Debeaujon et al., 2000, Appelhagen et al., 2014), while in legumes, naturally occurring pigmentation mutants have been reported to decrease testa thickness (Matus et al., 1993, Penmetsa et al., 2016), however there is less direct evidence for its role in permeability.

5.4.1 Pigmentation regulating permeability

In numerous legume species, phenolic content has been correlated with physical seed dormancy (Werker et al., 1979, Legesse and Powell, 1996, Souza and Marcos-Filho, 2001, Ross et al., 2010,

Freixas Coutin et al., 2017) which is strongly implicated by the Proanthocyanidin (PA) content (Zhou et al., 2010). Oxidation of PA causes brown pigmentation (Albert et al., 1997) and it is this oxidation process which has been linked to its impermeable properties (Marbach and Mayer, 1974, Moïse et al., 2005, Zhou et al., 2010). This was shown by normally dormant, wild *P. sativum var elatius* seeds becoming highly permeable when pigmentation expression was inhibited by the absence of oxygen (Marbach and Mayer, 1974). Imbibition rates in chickpea, French beans and pigeon pea, also showed rapid decrease in permeability with the increase of pigmentation expression (Legesse and Powell, 1996). How pigmentation causes hardseededness is not fully understood, however it is thought the crosslinking of Pas with other compounds form hydrophobic tannin – protein complexes (Oh et al., 1980, Girard et al., 2018) by non-covalent hydrophobic bonds may be responsible (Murray et al., 1994). Therefore, alterations to the phenolic content would preclude development of this hydrophobic layer.

An alternative mechanism how pigmentation might be regulating seed dormancy is the diverting of resources from the testa to the seed in non-pigmented lines. Sucrose can be converted to acetyl-coenzyme A (CoA) and malonyl-CoA, which are both precursor components to fatty acid (Baud et al., 2002, Baud et al., 2008) and flavonoid biosynthesis (Lepiniec et al., 2006). Sucrose acts as a major carbon source in the phenylalanine/flavonoid biosynthesis pathway, with around 20% of carbon allocated to the flavonoid biosynthesis pathway (Haslam, 1993). This constitutes a major competitive pathway for resources with accumulation of seed storage compounds (Baud et al., 2008, Lin et al., 2012, Li et al., 2018). The diversion of resources away from the testa would logically reduce fatty acid accumulation and composition, a component strongly linked permeability (Zeng et al., 2005, Shao et al., 2007). This is linked by the increased propensity for seed coat cracking (Ma et al., 2004a, Vu et al., 2014) with lines exhibiting reduced hydroxylated fatty acid content (Chai et al., 2016, Lashbrooke et al., 2016, Cechová et al., 2017, Gou et al., 2017). This is supported by the collective evidence from an experiment reducing pigmentation in navy beans causing reduced fatty acid content (Ross et al., 2010) and increased level of seed coat cracking from pigmentation mutants lines in soybean (Nicholas et al., 1993, Zabala and Vodkin, 2003) and peanut (Wan et al., 2016). This is again supported in pea transcriptome analysis between wild and domesticated accessions which showed a concurrent reduction in expression of certain flavonoid biosynthesis pathway genes and long chain fatty acid composition (Cechová et al., 2017, Hradilová et al., 2017). The targeting of fatty acid biosynthesis genes during domestication has been observed in numerous legume species (Zeng et al., 2005, Jang et al., 2015, Cechová et al., 2017, Hradilová et al., 2017) which could be linked to its seed coat permeability properties.

The mechanisms determining resource allocation are largely unknown, however flavonoid biosynthesis genes and two components of the MYB complex *WD40* and *MYB* (*TTG1* and *TT2)* were shown to direct resources to the seed coat by supressing seed storage accumulation (Baud et al., 2008, Chen et al., 2015, Li et al., 2018). As seed size is determined by the sucrose/glucose ratio (Weber et al., 1997) increased seed size in non-pigmented seeds (Debeaujon et al., 2000, Garcia et al., 2005, Chen et al., 2014, Chen et al., 2015), may be the consequence of increased partitioning of sucrose from the testa to the embryo. This is consistent with findings from Chapter 4 which mapped a seed weight QTL (*SW2*) near the Mendel *A* gene.

We were unable to directly test the effect of loss of *A* function mutant on permeability due to the lack of any induced *a* mutants, but two different induced anthocyanin-deficient mutants at the *A2* locus showed reduced testa thickness but no effect on permeability. *A2* encodes the *WD40* component of the *MBW* transcription factor that has a key role in the transcription of genes in the flavonoid pathway necessary for anthocyanin formation. Whilst it cannot be ruled out that Mendel *A* is part of a separate pathway working independently to *A2* or vice versa, these results strongly imply that *A* might be the causative gene for *TT2*, which is consistent with previous findings in other species (Debeaujon et al., 2001, Li et al., 2012, Chen et al., 2014, Qi et al., 2017).

A surprising result was that no effect to permeability was detected, since *WD40* ortholog *TTG1* plays a major role in allocating resources to seed coat as previously discussed (Baud et al., 2008, Chen et al., 2015, Li et al., 2018), and in the arabidopsis model was found to have the greatest effect on permeability among 20 flavonoid pathway genes (Debeaujon et al., 2000). One interpretation of this result is that the causative gene for dormancy loss in pea is closely linked but independent to Mendel *A*. This is supported by the presence of pigmented domesticated pea lines, which clearly show that the reduced dormancy associated with domestication does not require the loss of pigmentation (Smykal et al., 2014, Hradilová et al., 2017). A second search for potential candidate genes within the region around *A* revealed four potential candidates (Table 5.4). These genes encode components of part of the fatty acid biosynthesis pathway and could affect permeability by modulating hydroxylated fatty acid content. Alternatively, the background line used in the induced mutant experiment may have had a second mutation which potentially masked the effect of the *A2* mutant. This was supported by the wild type (JI2822) exhibiting high permeability rates (~1 to 2 days) like non-pigmented domesticated NGB5839 parental line. Therefore, both wild and induced mutant lines exhibited near free germinating phenotypes which as a result prevented the ability to make comparisons in permeability between wild and induced mutants. Consequently, Mendel *A* is still considered a strong candidate for regulating permeability, although this could not be conclusively shown. This is consistent

with reports of *bHLH* mutants exhibiting reduced hydroxylated fatty acid content in other separate species (Chen et al., 2014, Qi et al., 2017) therefore implying reduction in permeability (Shao et al., 2007).

*Table 5. 4: List of candidate genes for loss of seed dormancy between markers NEK4 and NAD. Potential genes were identified using candidate gene approach with M. truncatula reference genome. Candidate genes were searched in Tayeh et al. (2015) if markers were not found closest mapped marker was given which is denoted by an *.*

Gene	Position	PsCam
Seed Maturation	Medtr1g072090	PsCam026796, PsCam023421*
Acetyl-CoA carboxylase	Medtr1g071610	PsCam017622
Lipid transfer protein	Medtr1g071720	PsCam035394, PsCam033808*
Lipid transfer protein	Medtr1g071730	PsCam035394, PsCam033808*

5.4.2 Evolution of pigmentation loss in pea

As dormancy loss has been recognised as a critical domestication trait, it would be expected that its causative gene would be incorporated across the domesticated germplasm (Weeden, 2018) as found with *Dpo1* locus in pea responsible for indehiscent pods (Weeden et al., 2002, Weeden, 2007, Hradilová et al., 2017, Weeden, 2018). However, despite findings strongly implying Mendel *A* as responsible for dormancy loss, non-pigmented domesticated lines are known to exist (Holdsworth et al., 2017). This implies pigmentation loss occurred later and therefore implying dormancy loss evolved earlier via a separate mechanism. This is consistent with findings in chickpea where loss of testa pigmentation has occurred repeatedly within the domesticated germplasm to generate multiple "Kabuli" lineages (Penmetsa et al., 2016).

Alternatively, loss of dormancy may have occurred numerous times and pigmentation loss represents just one of these possible pathways. This would support the theory that pea was domesticated in multiple independent sites (Fuller et al., 2012). To investigate the evolution of pigmentation loss in pea, 431 accessions were examined by comparing phenotype to genetic diversity and subspecies level, using data from Holdsworth et al. (2017). As shown in Figure 5.11, genetic diversity of wild *P. sativum var elatius* and domesticated *P. sativum var sativum* showed distinct groupings likely driven by their independent evolutionary trajectories. A minor overlap was observed between wild and domesticated lines which likely represents the origin of domesticated pea. A non-pigmented wild line was identified near the cultivated accessions (highlighted in Figure 5.11 by "z"). This line (W6_15008) originates from Israel, part of the Fertile Crescent, and coincides with the suspected origin to the domesticated *P. sativum var sativum* (Zohary and Hopf, 1973, Lev-Yadun et al., 2000, Jing et al., 2010). This line could represent an early ancestor to the domesticated lines selected for its free germinating habit.

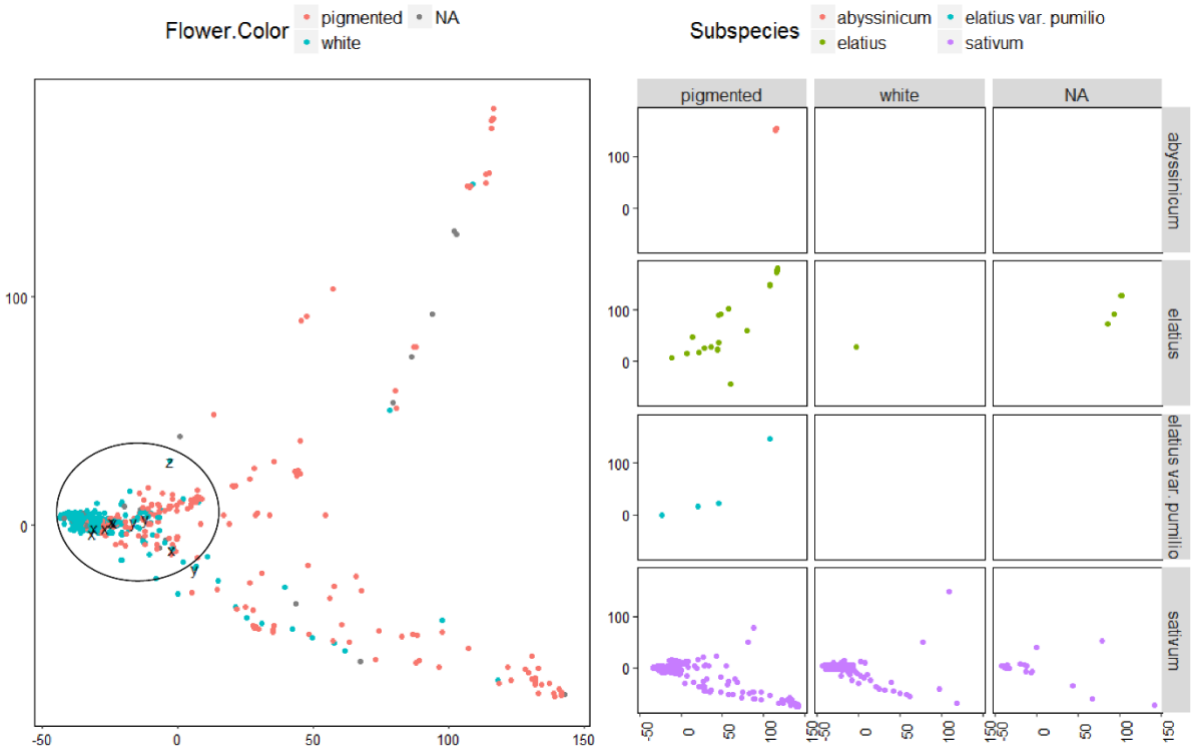


Figure 5. 11: Principal component analysis using data from Holdsworth et al. (2017) to compare genetic diversity in relation to pigmentation and species. Z = wild non-pigmented line, x and y represent infrequent loss of pigmentation alleles x = alternative bHLH loss of function allele, y = WD40 loss of function allele. Circle indicates major overlap between wild and domesticated lines.

In peas, there have been three independent mechanisms whereby pigmentation loss has reported (Hellens et al., 2010). Two resulted from independent mutation to the *bHLH* gene either by a misplice (*a*) or an INDEL (*INDEL*), while the third was due to the *WD40* gene (*a2*). These alleles were examined across 138 domesticated pea accessions which included 60 pigmented lines (*A*), Hellens et al. (2010) reported loss of pigmentation was dominated by the *a* allele accounting for 78 of the 88 lines, whereas *INDEL* and *a2* alleles represented only 7 and 3 lines respectively. The evolution of these accession was examined by cross referencing genetic diversity results from the Holdsworth et al. (2017) study. As shown in Figure 5.12, all lines regardless of mutation or pigmentation type were tightly clustered, indicating very similar genetic architecture. The non-pigmented and pigmented lines were not genetically distinct but showed significant overlap in the region previously been recognised to represent basal divergence of wild and domesticated lineages. Based on this evidence I propose pigmentation loss was likely incorporated during an early domestication event. Whether this is responsible for loss of dormancy or is closely linked cannot be conclusively shown, however it is clear that loss of seed dormancy has evolved separately to pigmentation.

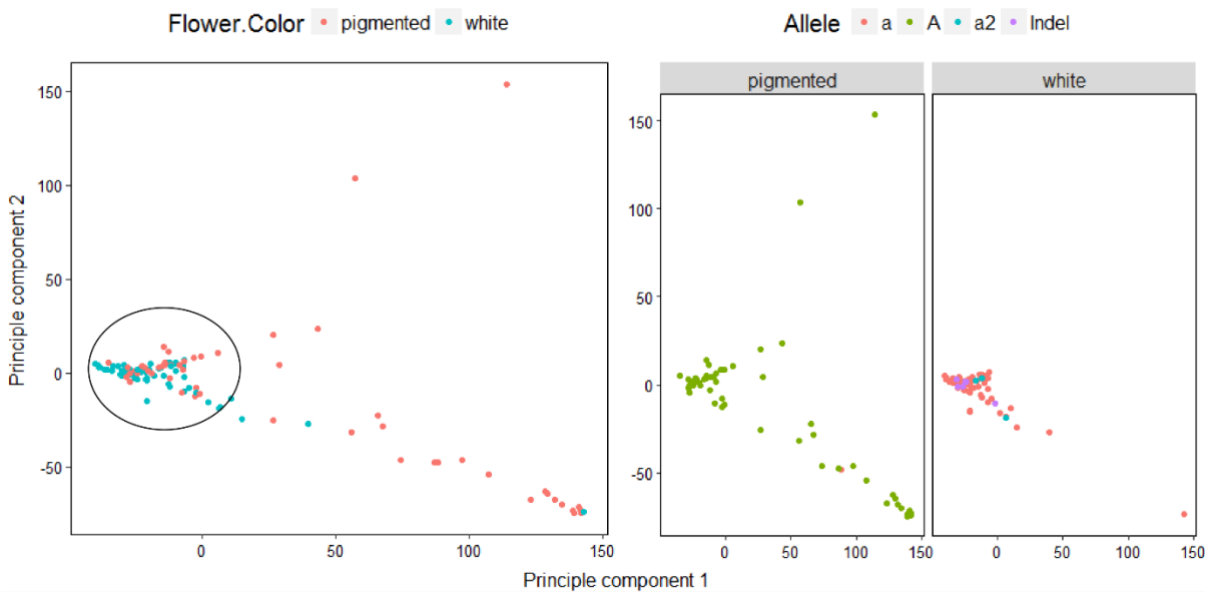


Figure 5. 12: Principle component analysis comparing genetic diversity of pigmented and non-pigmented lines analysed in (Hellens et al., 2010) using genetic diversity analysis data from Holdsworth et al. (2017). Pigmented lines are represented as a single allele "A", whereas non-pigmented lines are subdivided into three alleles "a", "INDEL" and "a2" with the former two representing independent mutations within the bHLH gene and the latter a mutation in the WD40 gene. Circle indicates significant overlap between wild and domesticated lines identified in Figure 5.11

5.5 Conclusion

Seed dormancy and testa thickness QTLs near the *A* locus on LGII have been validated using advanced generation segregating populations. These two traits showed strong co-segregation and are likely regulated by the same gene. A lower and a tentative upper QTL boundary were identified and calculated as a 4.4cM interval in relation to the Tayeh et al. (2015) linkage map, however validation of the upper QTL boundary requires further work. This can be achieved by advancing lines with specific recombination events which have been identified in the discussion.

Mendel *A* gene was recognised as a strong candidate for both QTLs, with loss of pigmentation associated with the thinning of the testa and increased permeability. Testa-thickness phenotypes in non-pigmented *a2* mutant lines, indirectly support the idea that *A* might be responsible for the TT2 QTL. However, its role in seed dormancy is less certain. Genetic analysis showed loss of pigmentation may have arisen from the early domestication phase, supporting its possible role in dormancy loss. However, the existence of pigmented domesticated lines indicates a separate mechanism has evolved. Whether the causative gene is closely linked to Mendel *A* or has evolved multiple times via independent pathways was not established in this study and future research is required. To validate Mendel *A*, induced mutant lines in a wild (impermeable and pigmented testa) background is required. Future efforts to analyse hydroxylated fatty acid content would be one way to explore the potential mechanism for regulating permeability proposed from this study. Furthermore, investigating the genetic control of seed dormancy in crosses of JI1794 with pigmented domesticated lines and

primitive land-races would provide further insight into whether *a* locus affecting permeability exists in this region when *A* is not segregating.

Chapter 6: Genetic dissection of dormancy related traits on LGVI

6.1 Introduction

The QTL analysis in Chapter 4 identified two significant QTLs for testa permeability; a major QTL on LGII (*Perm2*) and a minor QTL on LGVII (*Perm7*). In Chapter 5, *Perm2* was validated using advanced generation segregating populations which had a strong association with testa thickness and pigmentation. The smooth testa trait, which is near-universal in the domesticated germplasm did not appear to influence permeability, even though it has previously been closely mapped to a testa thickness QTL with testa thickness implicated in permeability (Smykal et al., 2014, Hradilová et al., 2017).

The rough testa phenotype typical of wild *Pisum* lines results from a granular surface texture caused by undulations within macrosclereid cell formation. A similar phenotype has been described in wild forms of numerous legume species including soybean (Otobe and Yoshioka, 2008, Otobe et al., 2015), chickpea (Toker, 2009), cowpea (Miao et al., 2001), common bean (Konzen and Tsai, 2014), lentil (personal observation), and pea (Blixt, 1972, Weeden and Wolko, 1990), but is commonly lost or reduced during domestication (Zohary and Hopf, 1973, Newell and Hymowitz, 1978, Lersten and Gunn, 1981, Plitmann and Kislev, 1989, Miao et al., 2001). In pea, smooth seededness is ubiquitous within both domesticated pea lineages; *P. sativum var sativum* and *P. sativum var abyssinicum* and is used as an archaeological indicator for domesticated material (Zohary and Hopf, 1973, Zaytseva et al., 2017). Whether the genetic basis for smooth seededness is similar in these two lineages is not entirely clear, although reported segregation in a cultivated x *P. sativum var sativum* population suggests independent genes (Weeden, 2007), its near-fixation in both suggests it has a strong connection to domestication.

Archaeological data has shown the transition from rough to smooth seededness occurred over 10,000 years ago (Zohary and Hopf, 1973). This corresponded with earliest evidence of cultivation of old world crops (Zohary et al., 2012). Because of this coordinated emergence of smooth-seededness with the timing of crop domestication, a link between testa roughness and physical seed dormancy has often been assumed. This has been tentatively supported by presence or absence of testa roughness correlating with dormant and non-dormant seeds in a small selection of pea genotypes (Cechová et al., 2017). Nevertheless, the causative region for testa roughness has not been well-defined in any legume species, and its potential role in regulating seed dormancy has not been carefully examined.

In pea, morphological mapping identified testa roughness as a single dominant monogenic trait known as *Gritty* (*GTY*)*;* (Marx, 1969). *GTY* has been previously mapped with relatively low resolution to LGVI (Weeden and Wolko, 1990, Weeden et al., 1998, Weeden, 2007) which is consistent with findings in Chapter 4. In addition to the QTL analysis in Chapter 4, a second novel minor QTL on LGVII was observed, therefore suggesting that expression of the trait is influenced by more than one gene. Interestingly both testa roughness QTLs mapped near a testa thickness QTL, as illustrated in Chapter 5 Figure 5.1. In both cases these had the same allelic direction with reduction in testa roughness correlated with reduced thickness, similar to previous reports (Zohary and Hopf, 1973, Lush and Evans, 1980, Plitmann and Kislev, 1989, Miao et al., 2001, Zaytseva et al., 2017). This suggests that testa roughness and thickness may be causally associated in some way, but it is surprising that a similar association is not seen between testa roughness or thickness and permeability. This raises questions about the validity of assumptions about the importance of testa thickness and roughness for domestication.

To address some of these questions, the work in this Chapter has investigated the classical testa roughness locus *GTY*, validating it and improving the resolution of its map position using advanced generation segregating populations, and examining its potential role in modulating seed dormancy by clarifying its genetic relationship with testa thickness and permeability.

6.2 Material and methods

6.2.1 Plant material and phenotype scoring

Populations segregating for the *GTY* region were obtained by advancing lines from the parental F_2 population of the RILs used for the QTL analysis in Chapter 4.

6.2.2 Phenotypic evaluation

Testa thickness was measured using a micrometer and permeability scored using the same technique described in Chapter 4. Testa roughness was measured as present or absent.

6.2.3 Genotyping

DNA extraction and genotyping were conducted using standard protocols described in Chapter 2. Details of markers used in this study are provided in Table 6.1.

*Table 6. 1: Details of markers used in this study. Marker position in pea is based on Tayeh et al. (2015) linkage map. * indicates when marker could not be mapped into pea linkage map and closest gene was used.*

Marker	Type	Sequences	Annealing Tm (°C)	*M. truncatula* position	PsCam	Position in pea
RUG5	HRM	F: CAGGTTCGTTCTGAATCTTCG R: CCTGCAGGAGTAAACGTGTG	52	Medtr6g012380	PsCam059118	40.6
COP13	HRM	F: ATAAAAGTTGATATGGGAGAAAGA R: CCAATGCAGGCACTCATA	58	Medtr6g023350	PsCam037990	46.1
APRL	HRM	F: TGGGATGCTTCCTATTGGTT R: TGAACATGGTCTGAAAATCTCAC	58	Medtr6g029240	PsCam051151	47.8*
FTa3	HRM	F: TTGTTCTTGGAGCTGTAATTGG R: CCTCAAATTTGGGTTACTAGGG	59	Medtr6g033040	PsCam005328	48.5*
MLO1	HRM	F: ACATTCCACCTGGCCTCAT R: TTGTGCATCATGTCCTGGAG	60	Medtr6g033330	PsCam045418	48.8
NT6083	HRM	F: CGTGTTTTTCTGAGTTGACTTCC R: TGTATACAGGGCAAACTCCTTG	53	Medtr6g034195	PsCam057485	49.1
MTIC153	HRM	F: TGCAACAAAAGAGGTATGAACTG R: TGGGTCGGTGAATTTTCTGT	58	-	PsCam057597	50.1
BFT	HRM	F: GGCCAATTTTGCTGATGACT R: TTTGACCACACTTGGTTCAACT	52	-	PsCam044479	-
GA20ox	HRM	F: GACCAACTTTTTAAGAAAAGCA R: TCTCCCATTTGAAAGAGCCTA	57	Medtr6g464620	PsCam054835	50.6*
RNAhel	HRM	F: GGGTTTGGTAGGTTTGGTAGAGG R: GCATGTGCTATTTTTCTTCACTC	57	Medtr6g056080	PsCam049373	-
CABB	HRM	F: AGGATCTTCTTGCCTGATGG R: CTTGCTTAGACCAAAAGGATCA	58	Medtr6g060175	PsCam050969	51.1
AGO1	HRM	F: TTACTCCCATGTCATCCTTGG R: CAAGCATTAAAGAACCAGCAAG	56	Medtr6g477980	PsCam043936	55.9
FULa	HRM	F: AACCTAGTAGCTCTCACCGTAA R: TTATATTATGGTGTTTGATTGATGA	52	Medtr2g461760	PsCam034575	56.8*
FVE	HRM	F: GGAGACTCCTCCGTCGCAGC R: TGTAGTTGCGCCTTTTTCG	57	Medtr2g039250	PsCam042915	63.8
PhyB	CAPs	F: AATCCCTTGAGTGGCATACG R: CAGCATGCAGAAGAGTGAGC	-	Medtr2g034040	PsCam035479	71.9

6.3 Results

6.3.1 Defining the Cluster 6 region

In Chapter 4, it was established that peak markers for testa thickness (*TT6*) and testa roughness (*GTY*) mapped within a 10cM region on LGVI. This can be seen in Figure 6.1, which has compiled testa structure and permeability data from chapter 4 in a detailed map spanning this region. The inclusion of anchor markers from the RIL population assists in mapping by comparing synteny in other model species. Results show that the peak DArT marker for testa thickness was situated between anchor markers *AGO1* and *FULa,* while testa roughness was mapped just above this position, between *CABB* and *AGO1*. No significant QTL was observed for permeability across this region.

As discussed in Chapter 3, previous studies have shown that a major chromosomal translocation event involving pea LGVI occurred during the divergence of the *Fabeae* tribe (Kaló et al., 2004, Tayeh et al., 2015), and can potentially be problematic in translational genomic studies. An analysis of the synteny of this region (referred to as Clst6), with *M. truncatula*, was performed in Chapter 3 but is presented again in Figure 6.2 at higher resolution. This shows that the Clst6 region spans the region affected by the translocation, and includes blocks of microsynteny with both Mt2 and Mt6. More detailed analysis indicates the overall level of synteny in this region is much lower than in most other regions, with additional insertions from chromosomes 1 to 6 as well as poor correspondence in marker order. It therefore appears this region has undergone major genomic reaarangements, and at this resolution *M. truncatula* is of limited use as reference for the corresponding region in pea.

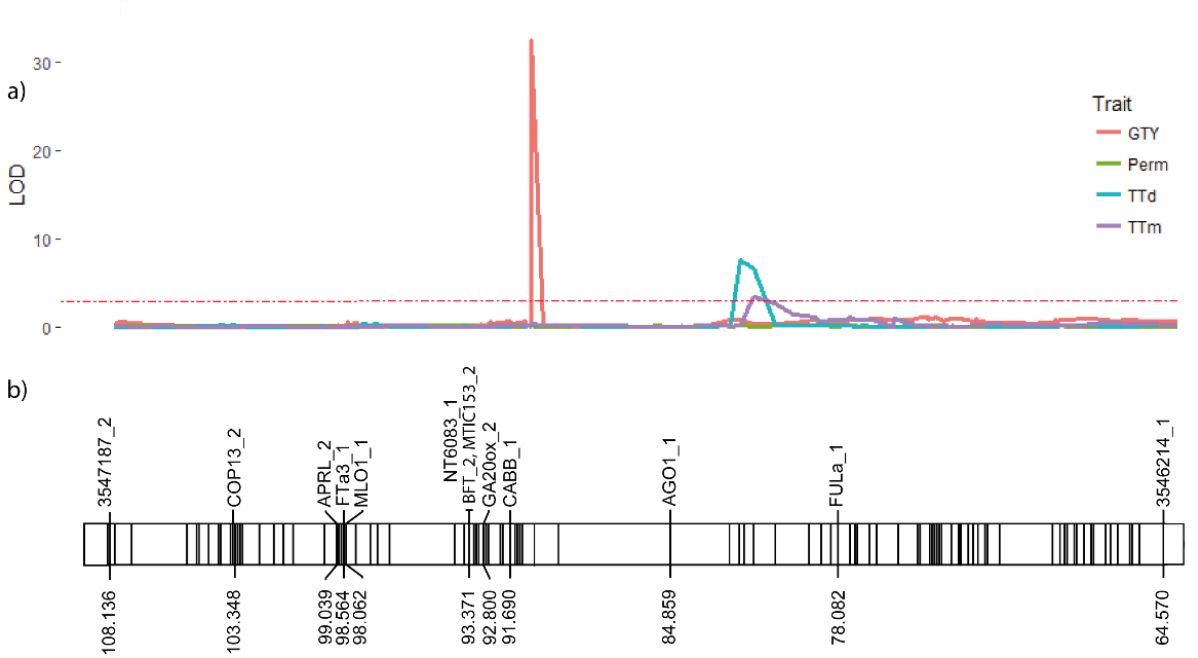


Figure 6. 1: Detailed QTL positions for testa roughness (GTY) and testa thickness measured using digital calliper (TT6d) and micrometer (TT6m) on LGVI between 64.570cM to 108.136cM. LOD score (a) and relative map positions (b) were calculated from QTL analysis on RIL population conducted in chapter 4. Each line black line represents marker, with anchor markers and their relative positions highlighted. Dashed red line illustrates LOD 3 threshold. Data extrapolated from Chapter 4.

As a possible alternative, synteny was compared with the more closely related species *L. culinaris*, as lentil chromosome 2 was shown in Chapter 3 to be collinear along its entire length with pea LGVI. Although this comparison was at a lower resolution than the Medicago comparison, since fewer pea markers mapped to the lentil genome in this region, synteny at this resolution was confirmed in Figure 6.2, supporting the idea that the major translocation event seen in the Medicago comparison must have occurred prior to the divergence of pea and lentil. Unfortunately, despite its closer phylogenetic affiliation, results show a lack of correspondance in marker order in this region, apparently indicating a complex assortment of inversions and reaarangements. While it is possible that significant structural changes have occurred following the divergence of *P. sativum L.* and *L. culinaris*, it is also possible that

these apparent differences reflect problems with the assembly of the lentil genome, in view of the the preliminary nature of the build. However, regardless of the real reason, this makes *L. culinaris* unsuitable as a reference for the corresponding region in pea, at present.

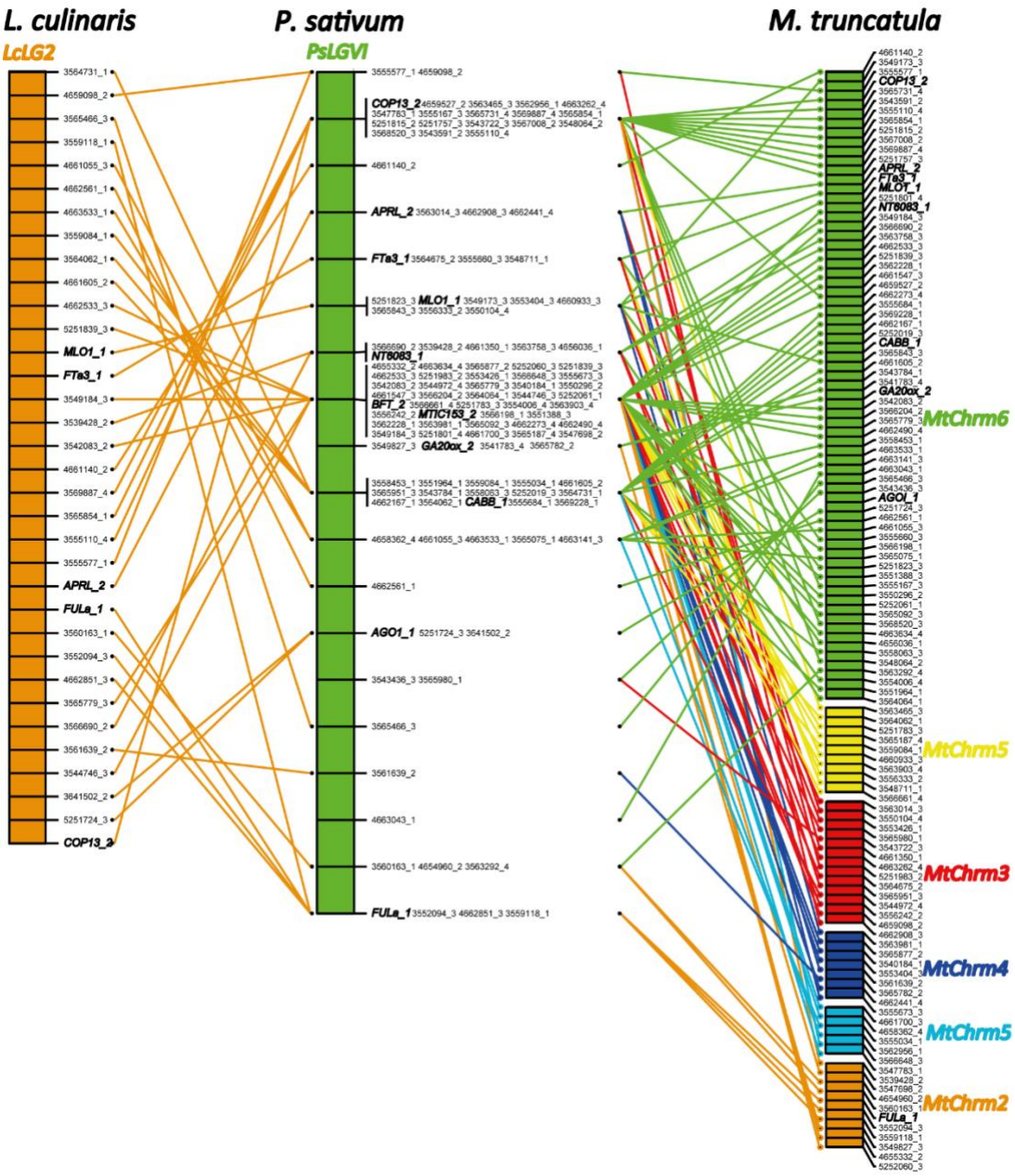


Figure 6. 2: Syntenic relationship of P. sativum Cluster 6 region with L. culinaris and M. truncatula genomes. Markers used in the development of the map were blasted into reference genome. Bold text signifies anchor markers. Colours represent different chromosomes/LG, Yellow = LG/Chrm1, Orange = LG/Chrm2, red =LG/chrm3 Dark blue = LG/Chrm4, Light blue = LG/chrm5 and Green = LG/Chrm6. Lines indicate homology.

6.3.2 Exploring the association between testa roughness and testa thickness

As previously discussed, testa roughness appeared to be closely associated with increased testa thickness. To visualise this association a correlation of these traits was analysed within the Clst6 region using phenotypic data collected in Chapter 4. Testa thickness and roughness scores were calculated for each marker by subtracting the average trait value given for all lines carrying the domesticated allele from the average trait value given for all lines carrying the wild allele. Positive values indicate wild trait dominance while negative values represent domesticated trait dominance. As shown in Figure 6.3, the scatter plot shows a positive correlation between testa thickness and roughness, as expected with closely linked traits.

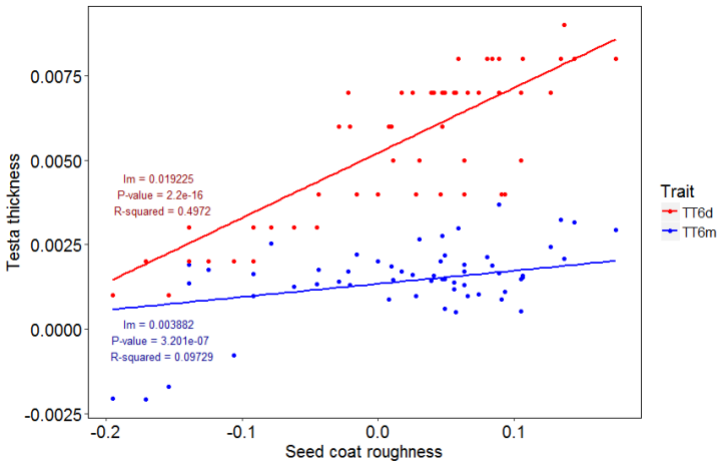


Figure 6. 3: Correlation between testa thickness and GTY using both testa thickness measuring techniques digital caliper (TTd) and micrometer (TTm).

To explore whether testa thickness and testa roughness were controlled by a single gene with pleiotropic effect or multiple genes with close linkage, I next attempted to evaluate QTL positions and intervals in greater detail. Figure 6.4 shows the relative positions of 45 markers across the region, with distances between markers represented by the cumulative number of recombinations (total 57) observed between markers in the RIL populations, rather than by cM. The results from Figure 6.4a confirm that testa thickness and testa roughness were strongly associated. This was seen by significant overlap in QTL region and corresponding allelic direction. Results also indicate the presence of two apparently distinct regions showing correlated effects on both traits. These regions will be referred to as *TT6a* and *TT6b*. This was unexpected, given that the original QTL analysis only identified a single QTL in this region.

As shown in Figure 6.4a, *TT6a* was putatively mapped between markers *CABB* to *FULa*, that were separated by 13 recombinations (recombinations 19 to 32). This region encompassed the two testa thickness peak markers (markers of interest 1 and 2) and testa roughness peak marker (marker of

interest 3) identified from the QTL analysis. The *TT6b* interval was putatively mapped between markers *APRL* and *COP13*, which were separated by seven recombination events (recombinations 3 to 10). This region did not include any testa thickness or testa roughness peak markers originally identified in the QTL analysis but did include the peak marker for testa roughness as indicated (marker of interest 4). The *TT6a* and *TT6b* regions were separated by 9 recombination events which occurred between markers *APRL* and *CABB*. It is currently unclear whether the increased testa thickness is promoted by testa roughness or whether *GTY* and *TT6* independently effect testa thickness.

a)

recombinations	Marker	Markers of interest**	Marker effect by trait			QTLs
			TTm	TTd	GTY	
0	3566093_2		0.170	0.005	0.063	
1	3563625_3		0.191	0.005	0.063	
3	4663171_4		0.243	0.007	0.127	*TT6b*
4	5251831_3	4	0.295	0.008	0.175	*TT6b*
5	COP13_2		0.324	0.008	0.134	*TT6b*
6	3547783_1		0.213	0.008	0.080	*TT6b*
9	4661140_2		0.219	0.007	0.049	*TT6b*
10	APRL_2		0.159	0.007	0.041	*TT6b*
11	4662908_3		0.119	0.007	0.056	
12	FTa3_1		0.144	0.007	0.039	
13	MLO1_1		0.089	0.006	0.008	
14	3540106_4		0.170	0.007	-0.022	
15	NT6083_1		0.140	0.006	-0.029	
16	MTIC153_2		0.170	0.007	0.017	
16	BFT_2		0.170	0.007	0.017	
17	3549827_3		0.139	0.007	0.056	
17	GA20ox_2		0.149	0.007	0.047	
19	CABB_1		0.061	0.007	0.049	*TT6a/GTY*
20	4658362_4		0.097	0.007	0.066	*TT6a/GTY*
21	4663141_3	3	0.102	0.007	0.074	*TT6a/GTY*
24	AGO1_1		0.154	0.008	0.106	*TT6a/GTY*
26	5251724_3		0.167	0.008	0.089	*TT6a/GTY*
27	3565980_1	2	0.208	0.009	0.137	*TT6a/GTY*
28	3565466_3		0.159	0.008	0.106	*TT6a/GTY*
29	3561639_2	1	0.148	0.007	0.105	*TT6a/GTY*
31	3560163_1		0.298	0.008	0.059	*TT6a/GTY*
32	FULa_1		0.368	0.007	0.089	*TT6a/GTY*
33	4663229_3		0.187	0.006	0.010	
35	3555837_1		0.098	0.004	0.028	
36	3566391_4		0.052	0.005	0.105	
37	3563708_4		0.090	0.004	0.091	
39	4659616_1		0.110	0.004	0.093	
40	3565520_3		0.131	0.004	0.063	
41	4663085_4		0.201	0.004	0.046	
43	3543728_3		0.133	0.003	-0.045	
44	3547129_3		0.126	0.003	-0.062	
45	3545667_3		0.097	0.002	-0.092	
47	3541373_3		0.136	0.002	-0.139	
48	3564941_4		0.175	0.002	-0.125	
49	3565051_4		0.254	0.003	-0.079	
51	4661025_1		0.191	0.003	-0.139	
52	4661698_3		0.162	0.003	-0.092	
54	3546214_1		0.175	0.004	-0.044	
56	4663291_1		-0.077	0.002	-0.106	
57	3551782_1		-0.169	0.001	-0.154	

b)

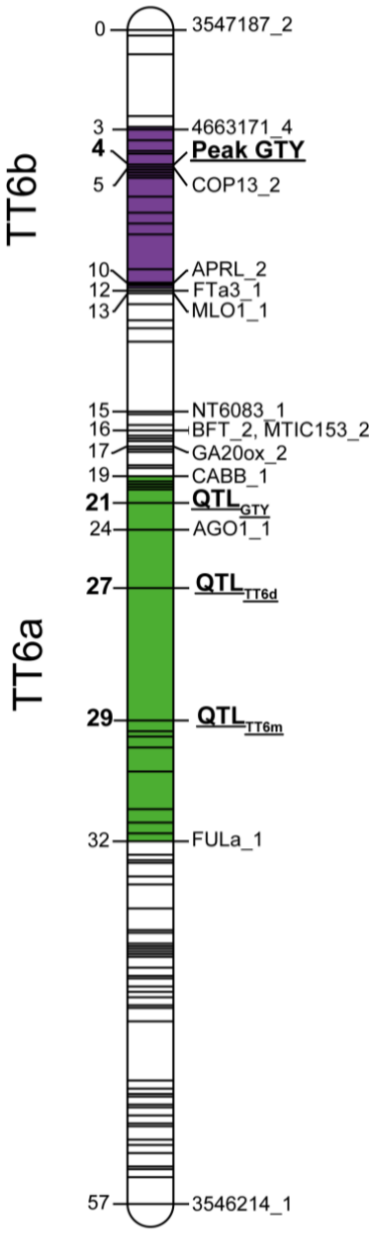


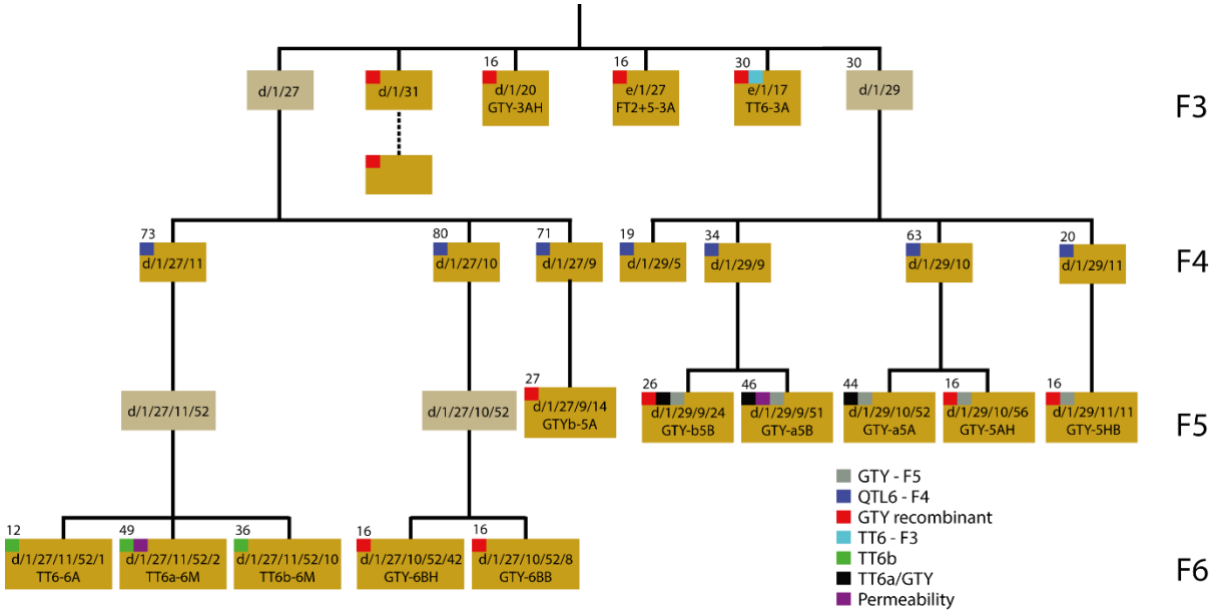
*Figure 6. 4: Informative markers identified in the RILs 20cM either side of peak TT6 QTLs. (a) Comparing the effects of testa thickness using micrometer (TTm) and digital caliper (TTd), and testa roughness in the RIL population across the Clst6 region. At each marker values for each trait is calculated by subtracting average domesticated value against average wild value, see supplementary Figure 6.1. QTL locations were determined by analysing significance in allelic direction and continuity of trait score across region. The significance of the wild (positive) or domesticated (negative) dominance was determined by trait value divergence from the 0 value. For each trait a colour range was used to illustrate changes in allelic direction from base level, with red showing increased allelic direction towards wild and blue showing increased direction towards domesticated. * Markers of interest 1 = highest variation in testa roughness between wild and domesticated alleles, 2 = peak marker testa roughness, 3 = peak marker testa thickness using digital caliper, 4 = peak marker for testa thickness using micrometer. (b) Illustrating the TT6a and TT6b across cluster 6 region, important markers and their respective positions shown.*

6.3.3 Development of advanced generation segregating populations

As detailed in Figure 6.5, segregating populations were developed to validate and refine the position of the testa thickness and testa roughness QTLs. This made use of the F_2 population originally used for RIL generation that had been partially genotyped in regions of interest. Populations were designed to segregate within the Clst6 region and selected wherever possible to be fixed in other genomic regions containing other relevant QTLs, as appropriate. This will be described in more detail in the relevant sections.

Identifying representative markers in the F_2 population for testa thickness (*TT*) and permeability (*Perm*) had previously been conducted in chapter 5. This included two flanking markers for *Perm2* and *TT2* (*LF* and *A*) a single marker for *Perm7* (*DUF*), a single marker for *TT7* (*SVPb*) and phenotypic marker for *TT5* (*Fs*). The Clst6 region was defined between markers *COP13* and *FULa*. No representative markers were available for *TT4*.

As part of a separate study investigating a flowering time QTL in this region, a large F_4 population (QTL6-F4) consisting of 360 individuals was also available, together with relevant genotyping data across the Clst6 region. With its large population size and dense genotyping in the Clst6 region, this population was used to identify and advance lines with a strategic recombination between *TT6a* and *TT6b* regions (between markers *APRL* and *CABB*) to be able to independently examine their phenotypic effect.



			QTLs							
Population	Parent population	Generation	TT2/ Perm2	TT5	TT6	TT7	Perm7	Population size	TT6 Segregating region	Function
FT2+5-3A	e/1/27	F_3	A	-	H	A	B	16	AGO1	3
GTY-3AH	d/1/20	F_3	A	-	H	A	B	16	BFT-CABB	3
TT6-3A	e/1/17	F_3	A	A	H	A	B	30	COP13-AGO1	3,4
QTL6-F4	1) d/1/27/9 2) d/1/27/10 3) d/1/27/11 4) d/1/29/5 5) d/1/29/9 6) d/1/29/10 7) d/1/29/11	F_4	A	H	H	H	B	360	MLO - FVE	2
GTY-a5A	d/1/29/10/52 (QTL6-F4 line 329)	F_5	A	A	H	B	B	44	RNAhel-FVE	1,6
GTY-a5B	d/1/27/9/51 (QTL6-F4 line 51)	F_5	A	B	H	B	B	46	BFT-FVE	1,6,7
GTY-b5A	d/1/29/9/14 (QTL6-F4 line 257)	F_5	A	A	H	B	B	27	FVE-AGO1	3
GTY-b5B	d/1/29/9/24 (QTL6-F4 line 267)	F_5	A	A	H	B	B	26	RNAhel-FVE	1,3,7
GTY-5AH	d/1/29/10/56 (QTL6-F4 line 333)	F_5	A	A	H	H	B	16	RNAhel-CABB	1,3
GTY-5HB	d/1/29/11/11 (QTL6-F4 line 351)	F_5	A	A	H	H	B	16	AGO1-FVE	1,3
GTY-6BH	d/1/29/10/52/42 (GTY-5B line42)	F_6	A	A	H	B	B	16	CABB	3
GTY-6BB	d/1/29/10/52/8 (GTY-5B line 8)	F_6	A	A	H	B	B	16	FULa-FVE	3
TT6-6A	d/1/27/11/52/1 (QTL6-F4 line 221)	F_6	A	B	H	A	B	12	COP13-FTa3	5
TT6-a6M	d/1/27/11/52/2 (QTL6-F4 line 221)	F_6	A	B	H	B	B	49	COP13-FTa3	5,7
TT6-b6M	d/1/27/11/52/10 (QTL6-F4 line 221)	F_6	A	B	H	H	B	36	COP13-FTa3	5

Figure 6. 5: Details of all advanced population analysed in this study a) Schematic representation illustrating the origin and development of populations, solid lines denote a direct link between populations/lines while dashed line denotes continued progression of lines of an unspecified level b) Table giving specific details about each advanced population including known genotype for testa thickness (TT) and permeability (Perm) QTLs and population size. No detail was given for TT4 as there is no available marker.

The Clst6 region was genotyped using 15 markers, detailed in Figure 6.6a. Marker order was established by the combined genotyping data from all populations used in this study, this is shown in Figure 6.6b. Overall there was strong consensus in marker order with no conflicts in positioning.

	Relative position	
Marker	Medicago (gene name and position)	L. culinaris (position)
RUG5	6g012380 chr6:3736016:3743000	LcChr2 171034321:171039143
COP13	6g023350 chr6 8148881:8139947	LcChr2 294473343:294480277
APRL	6g029240 chr6:10053541-10059871	LcChr2 190745791:190748615
FTa3	6g033040 chr6 10371673:10369989	LcChr2 117047987:117049228
MLO1	6g033330 chr6 10666895:10661399	LcChr2 104680595:104683674
NT6083	6g034195 chr6 11591273:11598330	LcChr2:169188046:169188967
MTIC153	-	-
BFT	-	-
GA20ox	6g464620 chr6 22554465:22552607	LcChr2 98741443:98743339
RNAhel	6g056080 chr6 20125418:20119486	LcChr2 95763839:95768883
CABB	6g060175 chr6 20668553:20672457	LcChr2 15464410:15467109
AGO1	6g477980 chr6 28760101:28768915	LcChr2 291549403:291554781
FULa	2g461760 chr2 25540748:25543882	LcChr2 200244722:200245929
FVE	2g039250 chr2 17196765:17189459	LcChr2 220530084:220538508
PhyB	2g034040 chr2 12993536:12986584	LcChr196736312:96736342

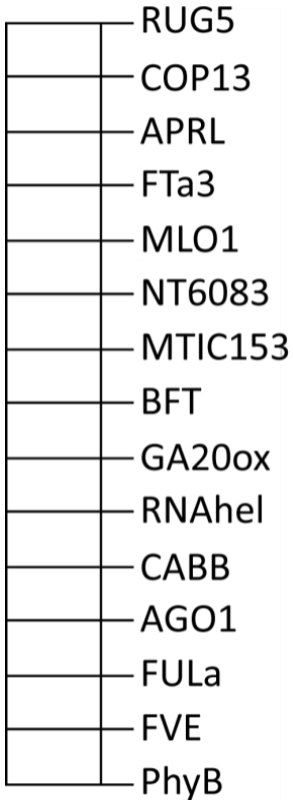


Figure 6. 6: Anchor markers within Cluster 6. Order based on combined genotyping from populations used in study.

6.3.4 Fine mapping GTY

To ascertain potential association of testa roughness with testa thickness, the *GTY* locus was initially fine mapped. Testa roughness has previously been recognised as a monogenic trait (Blixt, 1972, Weeden and Wolko, 1990) but the QTL analysis in Chapter 4, identified a second, minor QTL on LGVII when testa roughness was scored as a semi-quantitative trait. To examine the segregation of the major testa roughness QTL as a Mendelian locus, phenotypic data from the RIL population was converted from a quantitative to a qualitative present or absent score, by considering lines with trait values ≤2 to carry the domesticated ("smooth") allele, while lines with a value ≥ 2.5 were considered to carry the wild (rough) allele. In some cases, individuals with intermediate values were reassigned if their score conflicted with the genotype of neighbouring markers (see Supplementary Figure 6.2). A total of 131 the original 136 lines were phenotyped, with a segregation of 63 wild and 68 domesticated lines that did not deviate from the expected 1:1 ratio ($\chi 2=0.19$, $P=0.34$) thereby giving confidence to the conversion parameters.

When mapped according to this segregation, the major testa roughness locus (considered equivalent to *GTY*) was positioned close to the peak marker for the original QTL, in the *TT6a* interval between markers *FULa* and *CABB* (Figure 6.7). Further refinement of this position was precluded by apparently conflicting data. The recombination in Line SH137 was positioned *GTY* between *AGO1* and *3565062*/*FULa*, which conflicted with lines SH43 and SH69 which positioned it between *AGO1* and *CABB*. It is difficult to explain the conflicting positions of *GTY* by phenotyping errors because the conflicting lines: SH43, SH69 and SH137; had a strong smooth seeded phenotype according to the original RIL phenotyping data, with scores of 1, 1 and 1.5 respectively. Phenotyping anomalies were also observed in lines SH11, SH52 and SH126 where neighbouring markers suggested contrasting phenotypes. Lines SH11 and SH126 had intermediate testa roughness scores of 2 (B) and 3 (A) respectively, but genotypes across the region were A and B, respectively, suggesting that they were likely miscategorised, whereas line SH52 had a strong testa roughness phenotype originally scored as 4 (A) in the RIL population.

a)

Region	Recombination	marker	SH2	SH11	SH12	SH21	SH28	SH37	SH42	SH43	SH44	SH50	SH52	SH55	SH62	SH68	SH69	SH78	SH80	SH87	SH89	SH100	SH106	SH126	SH129	SH131	SH135	SH137
TT6b	5	COP13	B	A	B	B	B	A	B	B	A	A	B	A	A	A	B	A	B	A	B	A	B	B	B	B	B	B
	10	APRL	B	A	A	B	B	A	B	B	A	A	B	A	A	B	B	A	B	B	B	A	A	B	B	B	A	B
	12	FTa3	B	A	A	B	B	A	B	B	A	A	B	A	A	B	B	B	B	B	B	A	A	B	B	B	A	B
	13	MLO1	B	A	A	B	B	A	B	B	A	A	B	A	A	B	B	B	B	B	B	A	A	B	B	B	A	A
	14	3540106	B	A	A	A	B	A	B	B	A	A	B	A	A	B	B	B	B	B	B	A	A	B	B	B	A	A
	15	NT6083	B	A	A	A	B	A	B	B	A	A	B	A	A	B	B	B	B	B	B	A	A	B	B	A	A	A
	16	MTIC153	B	A	A	A	B	A	B	B	A	A	B	B	A	B	B	B	B	B	B	A	A	B	B	A	A	A
	16	BFT	B	A	A	A	B	A	B	B	A	A	B	B	A	B	B	B	B	B	B	A	A	B	B	A	A	A
	17	GA20ox	B	A	A	A	B	A	B	B	A	A	B	B	B	B	B	B	B	B	B	A	A	B	B	A	A	A
	19	RNAhel	B	A	A	A	B	A	.	.	A	A	B	B	B	B	B	B	B	B	B	A	A	B	B	A	A	A
TT6a	19	CABB	.	A	A	A	B	A	A	B	A	A	B	B	B	B	B	B	A	B	B	A	A	B	B	A	A	A
		GTY	A	A	A	A	B	A	A	B	A	A	B	B	B	B	B	B	A	B	A	A	B	B	B	A	A	B
	24	AGO1	A	A	A	A	B	A	A	A	A	A	B	B	B	B	A	B	A	B	.	A	B	B	B	A	A	A
		GTY	A	B	A	A	B	A	A	B	A	A	A	B	B	B	B	B	A	B	A	A	B	A	B	A	A	B
	28	3565466	A	A	A	A	A	A	A	A	A	A	B	B	B	B	A	B	A	B	A	A	B	B	A	A	A	B
	32	FULa	A	A	A	A	A	B	A	A	B	A	B	B	B	B	A	B	A	B	A	.	B	B	A	A	B	B
	33	4663229	A	A	A	A	A	B	A	A	B	B	B	B	B	B	A	B	A	B	A	A	B	B	A	A	B	B

b)

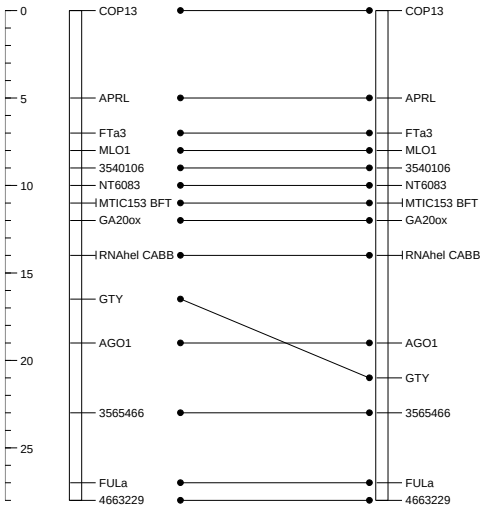


Figure 6. 7: Re-mapping GTY in RIL population based on neighbouring markers when converted to qualitative trait. (a) shows alternative map positions of GTY (Red) based on the genotype of neighbouring markers across RIL population. RIL lines showing conflicting positions of GTY are highlighted in colour orange = AGO1 and 3565466, yellow = CABB and AGO1, grey = miscellaneous. TT6a/GTY and TT6b regions highlighted in dark green and purple respectively. (b) shows alternative positioning of GTY based on recombinations.

6.3.4.1 Fine mapping GTY using advanced generation segregating populations

In an attempt to more accurately define the position of *GTY*, the presence or absence of *GTY* was evaluated on several advanced generation segregating populations, including the previously studied QTL6-F4 population described above. Unfortunately, not all individuals from this population were retained for seed production and it was therefore only possible to phenotype *GTY* in 47 of the original 360 lines (Figure 6.8a). However, to increase effective population size, four F_5 progenies were also

generated from selected F_4 individuals, taking advantage of specific recombination events to limit the segregating region (Figure 6.8b-e). This amounted to a total of 190 individuals which were scored for *GTY* and genotyped for seven markers (*AGO1*, *CABB*, *FTA3*, *FULa*, *FVE*, *PHYb* and *RNAhel*). Markers were ordered based on observed recombination events. Figure 6.8 showed that this was consistent across all five populations, and with the order determined in the RIL population (Figure 6.7).

Figure 6.8 shows that the F_5 family GTY-a5B (n = 46; derived from F_4 line 51) was only segregating between *FTa3* and *PHYB*; the family GTY-a5A (n = 44; derived from F_4 line 257) was only segregating below *RNAhel*, confirming the position of *CABB* below *RNAhel*. Family GTY-b5B (n = 26; derived from line no. 267) was only segregating below *CABB*. In these three cases, the F_5 results were consistent with those for the corresponding F_4 parent. The result from the third family in particular (Figure 6.8d) identified an additional recombination between *CABB* and *GTY*, confirming *CABB* as the upper flanking marker for the position of *GTY*, as indicated by RIL lines SH89 and SH137 (Figure 6.7).

The F_4 plant 329 had previously been genotyped as homozygous wild (A) for *AGO1* and heterozygous (H) for *RNAhel* (Figure 6.8a) but in its F_5 progeny (Figure 6.8e; family GTY-b5A, n = 27) *AGO1* was found to segregate, and *RNAhel* was found to be homozygous wild (A). In addition, there was no apparent segregation of *GTY* or of markers above including *CABB*, therefore this family provided an unambiguous confirmation for the location of *GTY* between *CABB* and AGO1. This conflicted with line GTY-a5A-8 (Figure 6.8c) which placed *GTY* between *AGO1* and *FULa*, however this was based on a single line rather than a population and therefore more likely to be artificial resulting from genotyping or phenotyping errors.

Across the 190 lines, 14 individuals could not be confidently phenotyped for *GTY*, indicated by the prefix “G?” in figure 6.8. From these 12 involved a situation where one or more flanking markers were heterozygous, which might suggest some degree of incomplete dominance, with heterozygous plants in some cases potentially exhibiting a weaker, intermediate phenotype appearing smooth (e.g. line GTY-a5B-39; Figure 6.8b). Nevertheless, this does not explain, where an individual was fixed for the domesticated genotype across entire region and therefore expected to be smooth seeded but were found to be *GTY* (e.g. line GTY-a5A-25; Figure 6.8b).

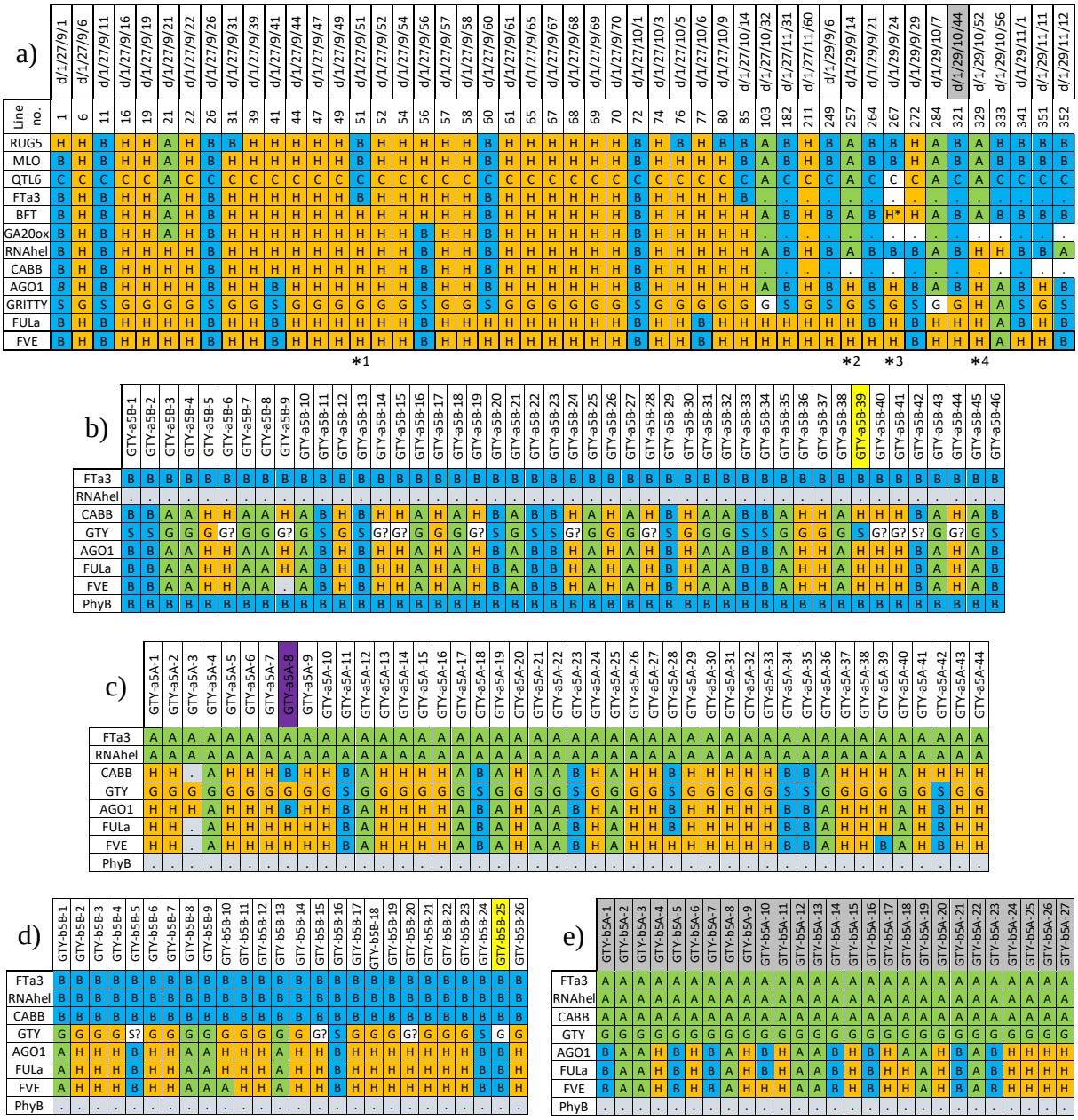


*Figure 6. 8: Fine mapping GTY in advanced segregating populations a) QTL6-F4, b) GTY-a5B, c) GTY-a5A, d) GTY-b5B and e) GTY-b5A. GTY was phenotyped as present (G) or absent (S), the suffix "?" indicating low phenotyping confidence. Colours indicate genotype with blue = homozygous domesticated, orange = heterozygous and Green = homozygous wild allele. Markers were ordered by best of fit, with Gty putatively mapped between AGO1 and CABB. Lines positioning GTY between AGO1 and CABB are highlighted in grey, while lines showing conflicting GTY positions are highlighted with yellow showing Gty outside FULa and CABB region and Purple outside AGO1 and CABB. Populations GTY-a5B, GTY a5A, GTY-b5B and GTY-b5A were derived from lines in QTL6-F4 population highlighted by *1-4 respectively.*

6.3.4.1.1 GTY recombinant populations

As previously described, one possible explanation for ambiguous mapping is errors in phenotyping and/or genotyping. As previously shown testa roughness appears to display quantitative expression influenced by modifier genes, and may also be inherited in an incompletely dominant manner, which may in some cases result in miscategorization. Another problem could potentially occur in genotyping of molecular markers, as scoring is not 100% accurate in every case and it is difficult to be confident in certain critical recombinants if based on genotyping of a single individual for a single marker. To explore whether such errors may have complicated the mapping of *GTY*, effort was also made to confirm critical single-plant genotypes by genotyping of their progeny. This was done in a total of ten cases, including six plants with an apparent recombination between *AGO1* and *CABB*, three plants with an apparent recombination between *FULa* and *AGO1,* and one plant with apparent recombination between both *FULa* and *AGO1* and *AGO1* and *CABB*. This included population GTY-6BB, derived from parental line GTY-a5A-8, which as previously discussed appeared to map *GTY* between *FULa* and *AGO1* (Figure 6.8c) as opposed to the general consensus *CABB* and *AGO1*. A total of eight markers spanning the *CABB-FULa* region were scored in all ten progenies.

Figure 6.9a presents the results from this analysis, which support the conclusion that *GTY* is located between *FULa* and *CABB*. This is shown with populations 1-4 indicating *FULa* as the lower QTL boundary while populations 7 to 9 indicating *CABB* as an upper QTL boundary. This has been consistent throughout this study, as shown in the summary of all relevant progenies examined (Figure 6.9a and b). Nevertheless, the position of *GTY* relative to *AGO1* has been more ambiguous, with analysis of the ten progenies failing to resolve apparent inconsistencies identified in their parent lines. Collectively, *GTY* was most confidently mapped between *CABB* and *AGO1*, (Populations 3 to 8), which throughout this study has now been mapped 13 times (as highlighted in grey in Figures 6.9a and 6.9b). This is most clearly seen in population 5 which is fixed for the domesticated *AGO1* allele and domesticated smooth seeded phenotype but segregating for *CABB*, and population 6 which is fixed for wild *CABB* allele and wild seed coat roughness phenotype but segregating for *AGO1*. This position is in contrast with population 10 which mapped *GTY* between *AGO1* and *FULa*, where throughout this study it has been independently mapped three times (as highlighted in red). While the cause for the alternative mapping of *GTY* is unknown*,* population 10 makes it difficult to explain this as genotyping or phenotyping error. One other possible explanation could be genomic rearrangements occurring within parental line. This is supported by findings in chapter 3 which recognised this region to be highly susceptible to rearrangements even between closely related species *L. culinaris*, however this is outside the scope of this study.

a)

Population		Parental population												
number	name	line	Population	Gen	FVE	FULa	AGO1	GRITTY	CABB	RNAhel	GA20ox	BFT	FTa3	MLO1
1	TT6-3A	e/1/17	F_2	F_3	A	A	H	H	H	H	H	.	.	H
2	-	d/1/31**	-	-	B	B	H	H	H	H	.	.	.	H
3	GTY-5AH	d/1/29/10/56	QTL6-F4 line 333	F_5	A	A	A	H	H	H	.	B	.	B
4	GTY-3AH	d/1/20	F_2	F_3	A	A	A	H	H	H	H	H	.	B
5	GTY-6BH	d/1/29/10/52/42	GTY- a5A-42	F_6	B	B	B	B	H	A	.	.	.	.
6	GTY-b5A	d/1/29/9/14	QTL6-F4 line 257	F_5	D	D	D	A	A	A	.	.	.	.
7	GTY-5HB	d/1/29/11/11	QTL6-F4 line 351	F_5	H	H	H	H	B	B	.	B	.	B
8	GTY-b5B	d/1/29/9/24	QTL6-F4 line 267	F_5	D	D	D	H	B	B	.	.	.	.
9	FT2+5-3A	e/1/27	F_2	F_3	H?	B	H	H	A	A	A	.	.	A
10	GTY-6BB	d/1/29/10/52/8	GTY- a5A-8	F_6	H	H	B	H	B	B	.	.		.

b)

Line	Gen	FVE	FULa	AGO1	GRITTY	CABB	RNAhel	GA20ox	BFT	FTa3	MLO1
QTL6-F4 - line 77	F_4	B	B	H	D	H	H	H	H	H	H
SH37	RIL	B	B	A	A	A	A	A	A	A	A
SH69	RIL	B	B	A	A	A	A	A	A	A	A
SH28	RIL	A	A	B	B	B	B	B	B	B	B
QTL6-F4- line 264	F_4	H	B	B	B	.	B	.	B	.	B
QTL6-F4- line 341	F_4	H	B	B	B	.	B	.	B	.	B
SH43	RIL	A	A	A	B	B	.	B	B	B	B
SH69	RIL	A	A	A	B	B	B	B	B	B	B
QTL6-F4- line 182	F_4	H	H	B	B	.	B	.	B	.	B
QTL6-F4- line 249	F_4	H	H	B	B	.	B	.	B	.	B
QTL6-F4- line 60	F_4	H	H	B	B	B	B	B	B	B	B
QTL6-F4- line 41	F_4	B	B	B	B	H	H	H	H	H	H
GTY-a5A-42	F_5	B	B	B	B	H	A	.	.	A	.
GTY-a5A-28	F_5	H	B	B	B	B	A	.	.	A	.
QTL6-F4- line 321	F_4	H	H	B	D	.	B	.	B	.	B
SH137	RIL	B	B	A	B	A	A	A	A	B	B

Figure 6. 9: Fine mapping Gty a) Position based from segregation patterns across population b) culmination of all important lines mapping Gty within the FULa-AGO1-CABB region. Lines or populations highlighted in grey map Gty between AGO1 and CABB, while lines or population in red map Gty between FULa and AGO1.

6.3.4.1.2 Identifying GTY using syntenic regions closely relating species

As previously shown in Figure 6.2, *GTY* is situated in a poor syntenic region between the pea map and physical maps of *M. truncatula and L. culinaris*, consequently restricting ability to design markers in this region due to its unpredictable mapping. In an attempt to find microsyntenic regions in which to design new markers, synteny was re-examined between newly refined *CABB-FULa* region using markers from the densely genotyped pea consensus map developed from the Tayeh et al. (2015) and included anchor markers *AGO1*, and *DOF6* for reference and orientation. As shown in Figure 6.10, synteny between *M. truncatula* reaffirms that the broader *GTY* position between *CABB* and *FULa* still represents a region of fragmented synteny that spans the Mt2 and Mt6 junction between markers *AGO1* and *DOF6*/*FULa*. Above *AGO1*, there is consistent synteny, but with some substantial disruptions of marker order - with a major inversion at the bottom of Mt6 relative to the pea map, between *AGO1* and *CABB,* as well as additional smaller inversions and rearrangements throughout this region. Comparative analysis between *L. culinaris* showed strong synteny with Lc2, but only weak collinearity, with substantial rearrangements in marker order. As discussed earlier, this may reflect

problems with assembly of the current (preliminary) lentil genome build. Based on the number of markers between *CABB* and *FULa*, the size of the *GTY* interval has expanded in both *M. truncatula* and *L. culinaris* genome, increasing from 11 markers in *P. sativum* to 47 in *M. truncatula* and 45 in *L. culinaris*. The combination of poor synteny and the expansion of the GTY region within the available reference genomes has precluded further refinement of the GTY loci. Collectively *GTY* can be confidently mapped to between *FULa* and *CABB*. Based on the high resolution transcriptome map developed by Tayeh et al. (2015) this 6.1cM region is inferred to include at least 48 genes which also mapped in *M. truncatula* with 37 genes occurring between CABB and the end of Mt6 (Supplementary Figure 6.3).

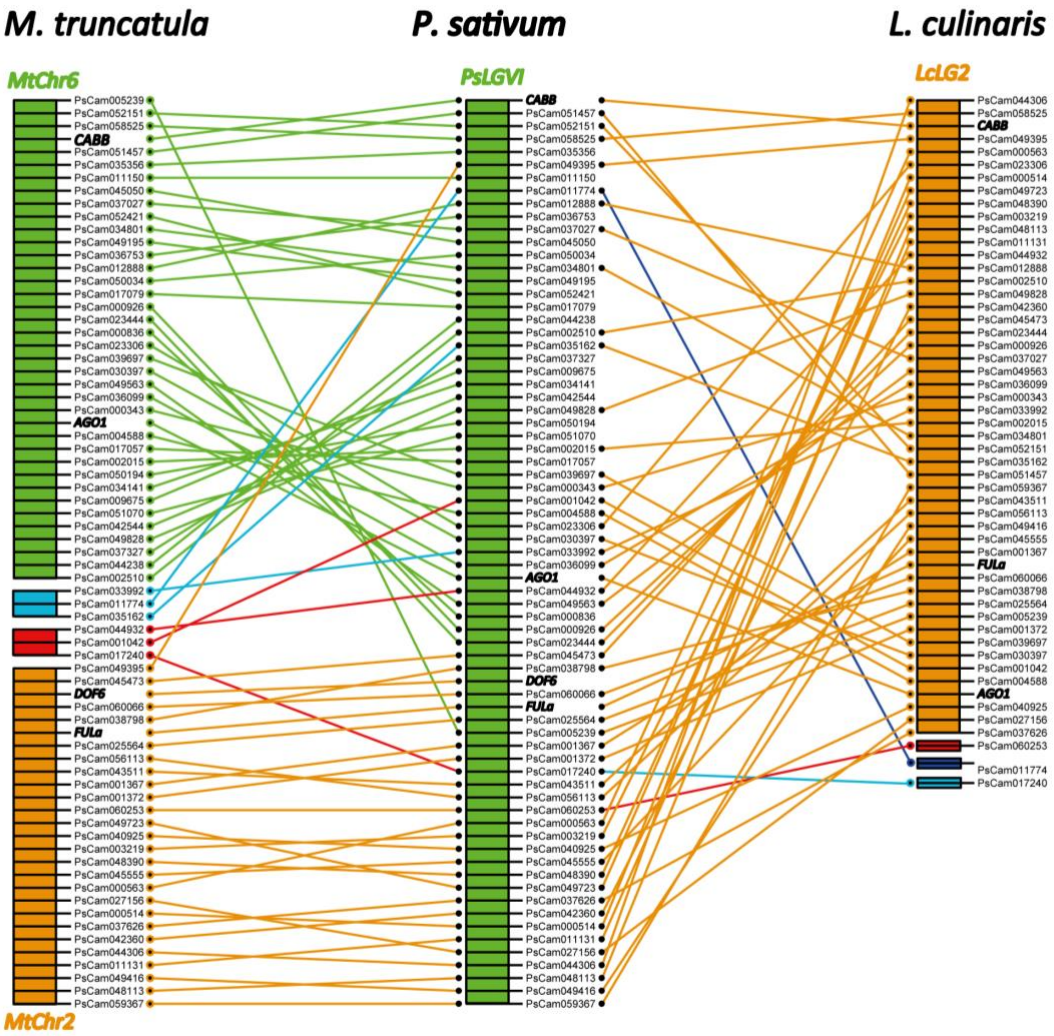


Figure 6. 10: synteny of TT6a region using PsCam markers used in Tayeh et al. (2015). Marker order determined by blasting Tayeh et al 2015 markers against reference genomes. Colours represent different chromosomes and LG, Orange = LG/Chrm2, red =LG/chrm3 Dark blue = LG/Chrm4, Light blue = LG/chrm5 and Green = LG/Chrm6. Lines indicate homology.

6.3.4.1.3 Validating testa thickness TT6a and TT6b

The second part of this chapter aims to validate both testa thickness QTLs identified in this region (*TT6a/GTY* and *TT6b*) and genetically clarify their relationship to testa roughness. Firstly, to validate

the existence of a testa thickness QTL an F_3 progeny (TT6-3A, *n*=30) was developed that was expected to segregate across the Clst6 region but not for additional testa thickness QTLs *TT2*, *TT5* and *TT7* (Table 6.2). Only the *TT4* genotype could not be determined since an appropriate marker had not been scored in the original F_2 population. Within the wider *TT6* region this progeny was segregating between *COP13* to *AGO1* but fixed between *FULa* and *FVE*. This meant while *TT6b* was clearly segregating, *TT6a* region could potentially be fixed. However, as the *TT6a* region is very close to *GTY*, which was segregating in the F_3 population, it was assumed this population was more likely also still segregating for *TT6a*.

Table 6. 2: Table detailing genetic background of TT and Perm QTLs in TT6-3A.

	QTL					TT6													
Population	TT2/perm	TT4	TT5	TT7	Perm7	FVE	FULa	AGO1	GTY	CABB	RNAhel	GA20ox	BFT	MTIC153	NT6083	MLO1	FTa3	APRL	COP13	Seed no.
TT6-3A	A	-	A	A	B	A	A	H	H	.	.	.	.	.	.	H	H	.	H	30
							TT6a/Gty											*TT6b*		

The TT6-3A population was genotyped using markers *COP13*, *FTa3*, *MLO1*, *GTY*, *AGO1*, *FULa* and *FVE* and phenotyped for testa thickness using a micrometer. As shown in Figure 6.11a, the genotyping results from the F_3 progeny confirmed it to be fixed wild for *FVE* and *FULa* and segregating between *AGO1* and *COP13.* In total three recombination events were identified, this included two between *GTY* and *FTa3* (lines 7 and 12) and one between *FTa3* and *MLO1* (line 19). No recombination events were identified between *FVE* and *FULa*, *AGO1* and *GTY* or *FTa3* and *COP13*.

With the exclusion of the additional QTLs, variation in testa thickness was much lower in TT6-3A population than the RILs. This ranged from 127.0 to 170.8µm (43.8µm) while the RILs ranged between 80.0 to 183.3µm (103.3µm) indicating the effect of many of the additional QTLs had been removed. As shown in Figure 6.11b and c, plants carrying the domesticated allele (B) had on average the thinnest testa (144.2µm), which was 4.5% and 4.4% thinner than wild (A) and heterozygous (H). This averaged at 151.0µm and 150.8µm respectively. This relatively minor variation in testa thickness corresponds to its small QTL effect observed in the RIL populations from chapter 4, which depending on measuring technique accounted for 12.3% of the 53.4% observed phenotypic variation using digital calliper or 5.5% of the 49.7% observed phenotypic variation using micrometer. The thinnest average testa was detected on markers *AGO1* and *GTY*. This corresponded to the peak testa thickness marker detected in the RIL population (Figure 6.4) and co-segregated with the mapped testa roughness QTL. Regardless, the Tukeys pairwise analysis did not detect a significant difference in testa thickness between genotypes and therefore could not validate *TT6*. This cannot be explained by conflicting

effects of *TT6a* and *TT6b* region as lines with recombination (highlighted in Figure 6.11c) did not appear to change the results. One possible explanation for this could be its genetic background, which potentially could still be segregating for *TT4*. Because *TT6* has a minor effect, results would be highly susceptible to the effect of additional QTLs, natural variation and phenotyping inconsistencies. By advancing populations further this would reduce the probability of additional effecting QTLs segregating within population by increasing homozygosity and greater population size would increase detectability.

a)

	TT6-3A-1	TT6-3A-2	TT6-3A-3	TT6-3A-4	TT6-3A-5	TT6-3A-6	TT6-3A-7	TT6-3A-8	TT6-3A-9	TT6-3A-10	TT6-3A-11	TT6-3A-12	TT6-3A-13	TT6-3A-14	TT6-3A-15	TT6-3A-16	TT6-3A-18	TT6-3A-19	TT6-3A-20	TT6-3A-21	TT6-3A-22	TT6-3A-23	TT6-3A-24	TT6-3A-25	TT6-3A-26	TT6-3A-27	TT6-3A-28	TT6-3A-29	TT6-3A-30
FVE	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
FULa	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
AGOI	H	H	A	H	H	H	H	H	B	H	B	H	H	A	H	H	H	A	H	H	A	B	H	H	B	H	A	B	H
GTY	D	D	D	D	D	D	D	D	B	D	B	D	D	D	D	D	D	D	D	D	D	B	D	D	B	D	D	B	D
FTa3	H	H	A	H	H	H	B	H	B	H	B	A	H	A	H	H	H	A	H	H	A	B	H	H	B	H	A	B	H
MLO1	H	H	A	H	H	H	B	H	B	H	B	.	H	A	H	H	H	H	H	H	A	B	H	H	B	H	A	B	H
COP13	H	H	A	H	H	H	B	H	B	H	B	A	H	A	H	H	H	H	H	H	A	B	H	H	B	H	A	B	H
Testa thickness	17.05	13.45	14.13	13.08	14.05	14.00	16.2	14.05	14.55	17.05	14.4	16.1	13.55	16.05	14.8	16.85	16.7	14.55	15.75	15.7	15.22	14.35	16.38	14.73	12.7	13.6	14.93	15.00	15.03

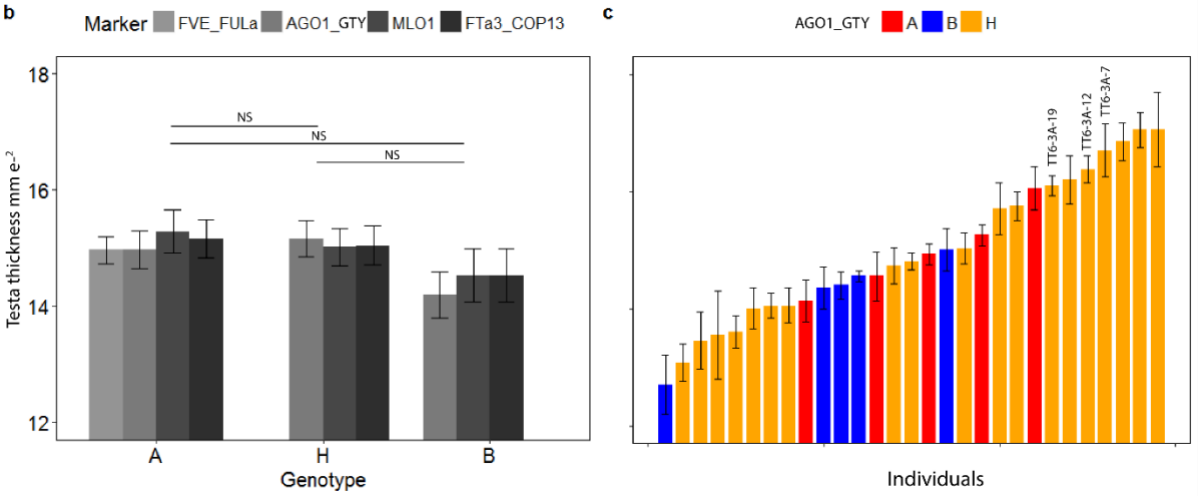


Figure 6. 11: Genotyping and Testa thickness scored across TT6-3A population. (a) Table showing genotyping scores (b) average testa thickness per genotype using 7 markers. FVE includes marker FULa and FTa3 includes marker COP13. (c) Range of testa thickness across population from thinnest to thickest, colours represent genotype based on AGO1 marker Blue = domesticated (B) red = Wild (A) and orange = heterozygous (H). Error bars calculated as standard error

6.3.4.1.4 Validating testa thickness QTL in TT6a/GTY region

To explore the association between testa thickness and roughness, testa thickness was also measured in the three F_5 progenies developed for the fine mapping of *GTY*. Importantly, these progenies all featured recombinations in the *CABB-MLO1* interval that fixed the genotype in the *TT6b* region but maintained segregation in the *TT6a* region. As shown in Table 6.3, all other testa thickness QTLs that could be genotyped were fixed, while the effect of *TT4* (which could not be genotyped due to lack of

a suitable marker in the F_2) was likely removed by the increased homozygosity associated with advancing the progeny from F_3 population (75%) to the F_5 (93.75%).

Table 6. 3: Development of TT6a/GTY population, markers in brackets denote known segregating regions. Asterix highlights potential genotyping error

		QTLs					TT6																
Population	parental line	TT2/Perm2	TT4	TT5	TT7	Perm7	PhyB	FVE	FULa	AGO1	GTY	CABB	RNAhel	GA20ox	BFT	MTIC153	NT6083	MLO1	FTAa3	APRL	COP13	RUG5	No. seeds
GTY-a5A	QTL6-F4 line 329	A	-	A	A	B	.	H	H	H	H	H	A	.	A	.	.	A	A	A	A	A	44
GTY-a5B	QTL6-F4 line 51	A	-	B	B	B	.	H	H	H	H	H	H	H	H	.	.	B	B	B	B	B	46
GTY-b5B	QTL6-F4 line 267	A	-	A	B	B	B	H	H	H	H	B	B		H*	.	.	B	B	.	.	B	26
									TT6a/Gty											*TT6b*			

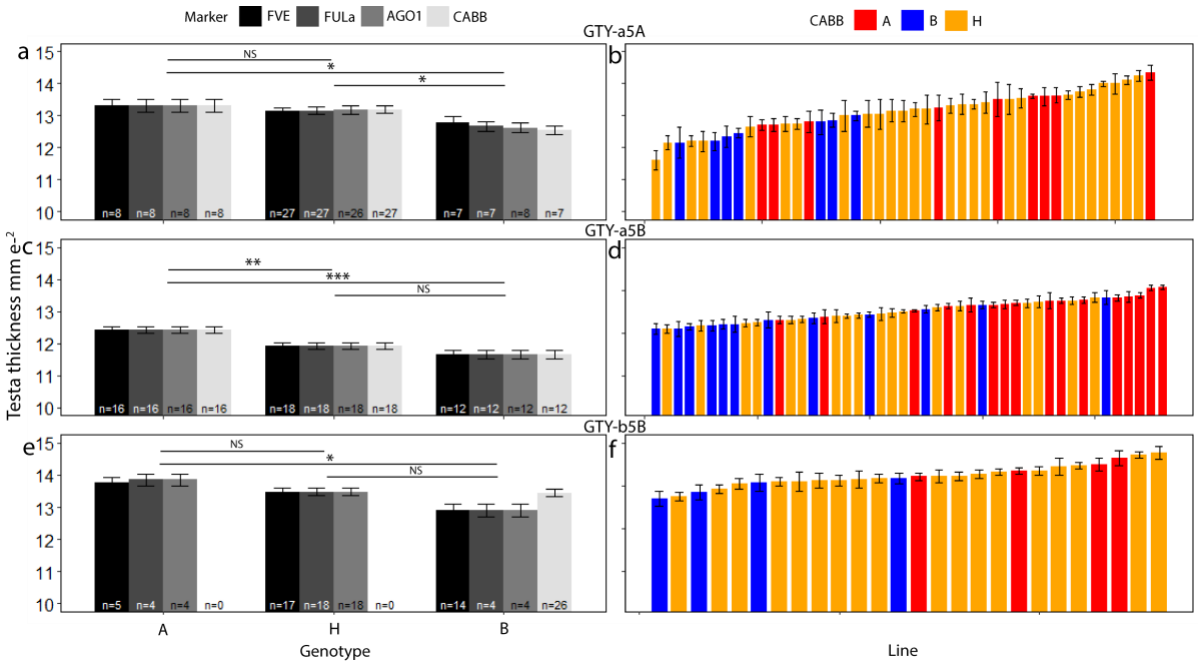
Testa thickness was measured using a micrometer and analysed against genotyping results described in Section 6.3.4.3 above (shown in Figure 6.8b-d). In all three populations the segregation ratio for wild:heterozygous:domesticated genotypes (A:H:B) did not conform to the expected 1:2:1. Population GTY-a5B exhibiting an unexpectedly high proportion of homozygous wild and domesticated genotypes but fewer than expected heterozygous individuals (16:12:18) and vice versa for GTY-a5B (4:18:4) and GTY-b5B (8:27:8). The range in testa thickness in these populations was 28, 20 and 22 µm respectively (data not shown). This was less than the 44 µm seen in the in the F_3 population and consistent with the removal of the effects of additional contributing QTLs such as *TT4* and *TT6b*.

Figure 6.12 shows that across all three populations, plants carrying domesticated alleles of *TT6a* (as inferred from the *AGO1* marker) exhibited the thinnest testas. This amounted to a 5-7% reduction in homozygous domesticated genotypes relative to wild which was determined to be significant by the Tukeys pairwise analysis, therefore validating the *TT6a*/*GTY* as a testa thickness QTL.

The position of this QTL can be definitively mapped between *PhyB* and *CABB,* based on the fixed recombination events in population GTY-b5B. This interval is estimated at 20.8cM based on Tayeh et al. (2015) linkage map. With regards to the thinning of the testa in population GTY-a5A, markers *AGO1* and *CABB* appeared most consistent to the mapping of QTL. As shown in Figure 6.12g this was supported by Tukeys pairwais analysis which showed these markers had the most significant difference in testa thickness between genotypes. Although this is based from a limited number of recombinations, this tentatively suggests testa roughness may be pleiotropically linked to increased testa thickness, as both testa thickness and testa roughness mapped closest to the *AGO1*-*CABB* region.

For mapping confidence, the distribution of plants carrying the domesticated, wild and heterozygous alleles of the *CABB* marker were compared against testa thickness range, as shown in Figure 6.12b, d

and f. Results showed plants carrying the domesticated alleles tended to congregate within the lower testa thickness range, while plants carrying the wild and heterozygous alleles were generally thicker. However, overlap in phenotypic range was observed in all genotypes which could indicate causative gene is mapped further away from the *CABB* marker, or the persistence of other genetic effects in the progeny. Regardless, it was considered more likely an issue in phenotyping accuracy, sample size and relatively small QTL effect. Collectively these results have shown that testa roughness and testa thickness co-segregated within a relatively narrow defined interval which strongly implies pleiotropic association. However, analysis of substantially larger population sizes in future may help to further clarify these effects and their genetic relationship.



g

	Tukeys Pairwise comparison (P-value)								
	FULa			AGO1			CABB		
population	A-H	A-B	B-H	A-H	A-B	B-H	A-H	A-B	B-H
GTY-a5A	0.8068	0.1173	0.1553	0.9000	0.0684	0.0534	0.8640	0.0433	0.0375
GTY-a5B	0.0033	0.0001	0.2944	0.0033	0.0001	0.2944	0.0033	0.0001	0.2944
GTY-b5B	0.3537	0.0265	0.0970	0.3537	0.0265	0.0970	-	-	-

*Figure 6. 12: Testa thickness measured across population (a-b) GTY-a5A (c-d) GTY-a5B and (e-f) GTY-b5B. Figures a, c and e show average testa thickness per genotype for markers FVE, FULa, AGO1 and CABB; Figures b, d and f shows distribution of testa thickness across population in relation to Marker CABB (Figures b and d) and AGO1 (Figure f). Significant differences in testa thickness between genotypes were calculated using Tukeys pairwise (NS >0.05, * P < 0.05, ** P<0.01, ***P<0.001) analysis, with results for each marker shown in Figure g. Significant differences are highlighted in yellow.*

6.3.4.1.5 Validating testa thickness QTL in TT6b region

As discussed above, a second potential testa thickness QTL, referred to as *TT6b*, was identified based on two apparent peaks in the strength of marker contributions across the larger LGVI region (Figure 6.4). This region appeared genetically separable from testa roughness which was mapped to the *TT6a*/*GTY* region. To explore the existence of a second testa thickness QTL acting independently to testa roughness, three advanced generation progenies fixed for testa roughness but segregating for

TT6b region were developed from the QTL6-F4 were. In view of the uncertainty associated with *GTY* map position and phenotyping, progenies were specifically selected to be unambiguously smooth seeded.

As shown in Table 6.4, these progenies were fixed for all additional testa thickness QTLs which could be genotyped, apart from population TT6-b6M which was still segregating *TT5*. Because *TT5* had previously been recognised as a minor QTL in chapter 4, accounting for 6.3% of the 53.4% observed phenotypic in the RILs when measured using digital caliper, while no significant effect was observed when using the micrometer, this was considered not to be a major complication. As previously discussed, *TT4* could not be accounted for since no suitable marker had been genotyped in the F_2, but was most likely fixed by the F_6 population, which would be expected to show 96.88% homozygosity).

*Table 6. 4: Development of TT6b population. Recombinations segregating based on values defined in Figure 6.4a. * TT6b and TT6a/GTY are separated by recombinations 10 and 17. To isolate regions recombination must occur between these recombination positions*

		QTLs					TT6														
Population	Parent population	TT2/Perm2	TT4	TT5	TT7	Perm7	FVE	FULa	AGO1	GTY	CABB	RNAhel	GA20ox	BFT	MTIC153	NT6083	MLO1	FTAa3	APRL	COP13	RUG5	No. seeds
TT6-6A	d/1/27/11/52/1	A	-	B	A	B	B	B	B	B	B	B	B	B	.	H	H	H	H	H	.	12
TT6-a6M	d/1/27/11/52/2	A	-	B	B	B	B	B	B	B	B	B	B	B	.	H	H	H	H	H	.	49
TT6-b6M	d/1/27/11/52/10	A	-	B	H	B	B	B	B	B	B	B	B	B	.	H	H	H	H	H	.	36
								TT6a/Gty											*TT6b*		

Plants were genotyped using markers *GA20ox*, *BFT*, *MTIC153*, *NT6083*, *MLO1*, *FTa3*, *APRL* and *COP13* and testa thickness was measured using a micrometer. As shown in Figure 6.13, genotyping of *GA20ox* and *BFT* confirmed populations were fixed between across the *TT6a*/*GTY* region. The inclusion of the *MTIC153* marker which was also fixed for the domesticated genotype provided further refinement of this region. A total of four recombination events were identified across all populations. Three recombinations occurred between *APRL* and *COP13* (TT6-b6M-33, TT6-a6M-3 and TT6-a6M-38) and one recombination between *NT6083* and *MLO1* (TT6-a6M-19). Distortion of the expected 1:2:1 ratio (A:H:B) was observed in all three populations. with a ratio that more closely resembled a 3:4:1, with an overrepresentation of homozygous wild genotype, as shown by 4:6:2, 18:25:6 and 12:19:5 in populations TT6-6A TT6-a6M and TT6-b6M respectively. Across each population variation between thickest and thinnest testa ranged from 20 to 40 µm, similar to the *TT6a*/*GTY* populations (20-28µm).

	GA20ox	BFT	MTIC153	NT6083	MLO1	FTa3	APRL	COP13	Testa thickness
TT6-6A-1	B	B	B	A	A	A	A	A	12.35
TT6-6A-2	B	B	B	A	A	A	A	A	12.10
TT6-6A-3	B	B	B	A	A	A	A	A	11.65
TT6-6A-4	B	B	B	H	H	H	H	H	10.85
TT6-6A-5	B	B	B	B	B	B	B	B	11.00
TT6-6A-6	B	B	B	A	A	A	A	A	11.35
TT6-6A-7	B	B	B	H	-	H	H	H	11.35
TT6-6A-8	B	B	B	H	H	H	H	H	11.05
TT6-6A-9	B	B	B	H	-	H	H	H	11.93
TT6-6A-10	B	B	B	B	B	B	B	B	11.40
TT6-6A-11	B	B	B	H	-	H	H	H	11.15
TT6-6A-12	B	B	B	H	H	H	H	H	12.89

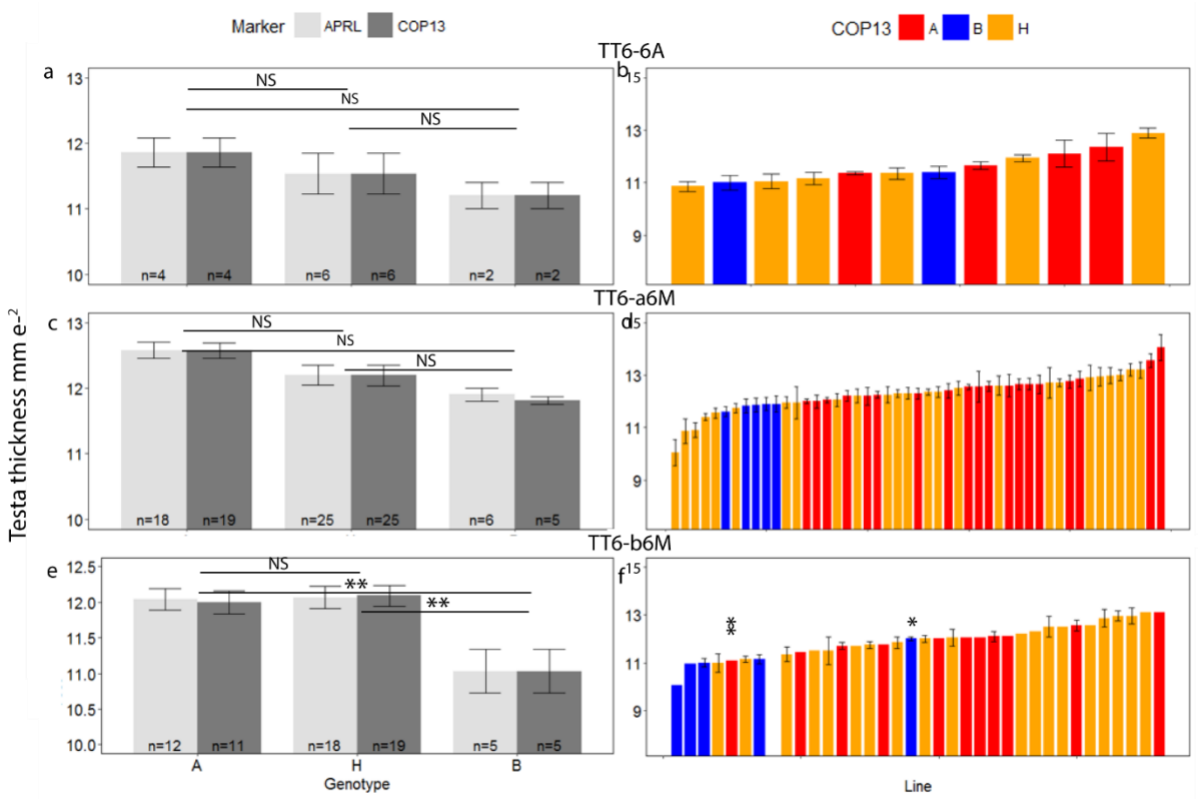
	GA20ox	BFT	MTIC153	NT6083	MLO1	FTa3	APRL	COP13	Testa thickness
TT6-b6M-1	B	B	B	H	H	H	H	H	12.50
TT6-b6M-2	B	B	B	H	H	H	H	H	11.15
TT6-b6M-3	B	B	B	A	A	A	A	A	12.10
TT6-b6M-4	B	B	B	H	H	H	H	H	11.50
TT6-b6M-5	B	B	B	H	H	H	H	H	11.35
TT6-b6M-6	B	B	B	B	B	B	B	B	11.15
TT6-b6M-7	B	B	B	B	B	B	B	B	11.00
TT6-b6M-8	B	B	B	B	B	B	B	B	12.00
TT6-b6M-9	B	B	B	H	H	H	H	H	11.50
TT6-b6M-10	B	B	B	H	H	H	H	H	11.00
TT6-b6M-11	B	B	B	H	H	H	H	H	12.00
TT6-b6M-12	B	B	B	H	H	H	H	H	11.85
TT6-b6M-13	B	B	B	H	H	H	H	H	12.95
TT6-b6M-14	B	B	B	-	-	H	-	-	11.30
TT6-b6M-15	B	B	B	H	H	H	H	H	12.05
TT6-b6M-16	B	B	B	H	H	H	H	H	12.95
TT6-b6M-17	B	B	B	A	A	A	A	A	11.70
TT6-b6M-18	B	B	B	H	H	H	H	H	11.75
TT6-b6M-19	B	B	B	A	A	A	A	A	12.55
TT6-b6M-20	B	B	B	H	H	H	H	H	12.85
TT6-b6M-21	B	B	B	A	A	A	A	A	11.75
TT6-b6M-22	B	B	B	A	A	A	A	A	12.05
TT6-b6M-23	B	B	B	B	B	B	B	B	10.95
TT6-b6M-24	B	B	B	A	A	A	A	A	12.00
TT6-b6M-25	B	B	B	H	H	H	H	H	11.70
TT6-b6M-26	B	B	B	A	A	A	A	A	12.10
TT6-b6M-27	B	B	B	H	H	H	H	H	12.20
TT6-b6M-28	B	B	B	H	H	H	H	H	12.50
TT6-b6M-29	B	B	B	A	A	A	A	A	12.05
TT6-b6M-30	B	B	B	B	B	B	B	B	10.06
TT6-b6M-31	B	B	B	H	H	H	H	H	12.30
TT6-b6M-32	B	B	B	A	A	A	A	A	11.45
TT6-b6M-33	B	B	B	A	A	A	A	H	12.55
TT6-b6M-34	B	B	B	A	A	A	A	A	11.10
TT6-b6M-35	B	B	B	H	H	H	H	H	13.10
TT6-b6M-36	B	B	B	A	A	A	A	A	13.10

	GA20ox	BFT	MTIC153	NT6083	MLO1	FTa3	APRL	COP13	Testa thickness
TT6-a6M-1	B	B	B	A	A	A	A	A	12.00
TT6-a6M-2	B	B	B	B	B	B	B	B	11.83
TT6-a6M-3	B	B	B	H	H	H	H	A	12.40
TT6-a6M-4	B	B	B	H	H	H	H	H	12.30
TT6-a6M-5	B	B	B	A	A	A	A	A	13.55
TT6-a6M-6	B	B	B	H	H	H	H	H	12.98
TT6-a6M-7	B	B	B	A	A	A	A	A	12.65
TT6-a6M-8	B	B	B	A	A	A	A	A	12.58
TT6-a6M-9	B	B	B	A	A	A	A	A	12.20
TT6-a6M-10	B	B	B	H	H	H	H	H	10.05
TT6-a6M-11	B	B	B	A	A	A	A	A	12.25
TT6-a6M-12	B	B	B	A	A	A	A	A	12.75
TT6-a6M-13	B	B	B	B	B	B	B	B	11.90
TT6-a6M-14	B	B	B	A	A	A	A	A	12.55
TT6-a6M-15	B	B	B	H	H	H	H	H	12.60
TT6-a6M-16	B	B	B	A	A	A	A	A	14.05
TT6-a6M-17	B	B	B	H	H	H	H	H	12.70
TT6-a6M-18	B	B	B	H	H	H	H	H	12.30
TT6-a6M-19	B	B	B	H	A	A	A	A	12.00
TT6-a6M-20	B	B	B	B	B	B	B	B	11.90
TT6-a6M-21	B	B	B	H	H	H	H	H	13.20
TT6-a6M-22	B	B	B	H	H	H	H	H	13.20
TT6-a6M-23	B	B	B	H	H	H	H	H	11.40
TT6-a6M-24	B	B	B	A	A	A	A	A	12.85
TT6-a6M-25	B	B	B	H	H	H	H	H	13.00
TT6-a6M-26	B	B	B	A	A	A	A	A	12.05
TT6-a6M-27	B	B	B	H	H	H	H	H	12.05
TT6-a6M-28	B	B	B	H	H	H	H	H	12.20
TT6-a6M-29	B	B	B	H	H	H	H	H	11.55
TT6-a6M-30	B	B	B	H	H	H	H	H	12.90
TT6-a6M-31	B	B	B	H	H	H	H	H	12.70
TT6-a6M-32	B	B	B	A	A	A	A	A	12.65
TT6-a6M-33	B	B	B	H	H	H	H	H	11.95
TT6-a6M-34	B	B	B	A	A	A	A	A	12.65
TT6-a6M-35	B	B	B	H	H	H	H	H	12.95
TT6-a6M-36	B	B	B	H	H	H	H	H	10.85
TT6-a6M-37	B	B	B	A	A	A	A	A	12.60
TT6-a6M-38	B	B	B	B	B	B	B	H	12.35
TT6-a6M-39	B	B	B	H	H	H	H	H	12.25
TT6-a6M-40	B	B	B	H	H	H	H	H	12.35
TT6-a6M-41	B	B	B	H	H	H	H	H	11.75
TT6-a6M-42	B	B	B	A	A	A	A	A	12.55
TT6-a6M-43	B	B	B	H	H	H	H	H	12.50
TT6-a6M-44	B	B	B	A	A	A	A	A	12.20
TT6-a6M-45	B	B	B	H	H	H	H	H	11.95
TT6-a6M-46	B	B	B	B	B	B	B	B	11.60
TT6-a6M-47	B	B	B	A	A	A	A	A	12.30
TT6-a6M-48	B	B	B	H	H	H	H	H	10.90
TT6-a6M-49	B	B	B	B	B	B	B	B	11.85

Figure 6. 13: Genotyping and testa thickness result across three advanced segregating populations for TT6b region. Colours represent genotypes green = wild (A), blue = domesticated (B) and orange = heterozygous (H). Unknown genotypes highlighted with a "- ". Testa thickness given as an average.

As shown in Figure 6.14, plants carrying the domesticated allele exhibited on average the thinnest testa across all three populations. This was calculated at 5.6%, 6.0% and 8.0% thinner than plants carrying the wild allele. A Tukey's pairwise analysis showed a significant difference in thickness between plants carrying wild and domesticated alleles for TT6b-b6M while TT6-a6M showed a strong trend but marginally outside significant threshold ($p < 0.05$). TT6-6A did not show a significant difference but this is likely due to its restricted population size. This collective evidence appears to validate a distinct testa thickness QTL within the *TT6b* region that is independent from testa roughness.

The marker *MTIC153* provides a lower QTL boundary as it was fixed across all three populations, but no clear upper QTL boundary could be defined. This is seen in the Tukeys pairwise analysis (Figure 6.14g) with markers *APRL* and *COP13* having similar P-values, resulting from the limited number of recombination events (TT6-b6M-33, TT6-a6M-3 and TT6-a6M-38). Based on results from the RILs (Figure 6.4) *COP13* showed greatest variation in testa thickness between plants carrying wild and domesticated alleles, therefore for mapping confidence, the distribution of plants carrying domesticated, wild and heterozygous *COP13* alleles were compared against testa thickness range. As shown in Figure 6.14b, d and f, results demonstrate strong clustering of the domesticated allele within the lower testa thickness range, providing a further illustration that *TT6b* is strongly linked to the *COP13* region. However, certain lines showed a phenotype which was not distinct from that of other individuals sharing the same genotype. This was seen for lines TT6-6A-10 and TT6-6M-8 which carried a domesticated allele but exhibited a thicker than expected testa and line TT6-6M-34 which carried a wild allele but had a thinner testa than expected (highlighted in Figure 6.14b and f). This could potentially reflect recombination between the causal *TT6b* gene and the *COP13* marker, given the possibility that *TT6b* might sit above rather than below *COP13* (Figure 6.4). To further refine this region, future work should include genotyping for more markers outside the *COP13* region, and a substantially increased population size to improve the detection and discrimination among marker effects.



g

	Tukeys Pairwise comparison (P value)					
	APRL			COP13		
population	A-H	A-B	B-H	A-H	A-B	B-H
TT6-6A	0.711	0.474	0.794	0.711	0.474	0.794
TT6-a6M	0.155	0.0832	0.585	0.154	0.064	0.46
TT6b-b6M	0.993	0.011	0.006	0.909	0.017	0.005

*Figure 6. 14: Effects of testa thickness in TT6b region in three populations; population TT6-6A (a and b), TT6-a6M (c and d) and TT6b-b6M (e and f). Figures a,c and d show average testa thickness across population and Figures b, d and e show segregation pattern of TT6b based on testa thickness phenotype using marker COP13. * TT6b-b6M-8, **TT6b-b6M-34. Error bars are calculated using standard error. Significant differences in testa thickness between genotypes was calculated using Tukeys pairwise (NS >0.05, * P < 0.05, ** P<0.01, ***P<0.001) analysis, with results for each marker shown in Figure g. Significant differences are highlighted in yellow.*

6.3.4.2 Effect of Clst6 to Permeability

The third aim of this chapter was to further examine whether variation for testa roughness and testa thickness on LGVI had any association with permeability. The QTL analysis in chapter 4 did not detect any effect of the Clst6 region on permeability, but it was considered important to confirm this in the advanced generation segregating material. Suitable populations were selected that had fixed genotypes at markers for the two known permeability QTLs (*Perm2* and *Perm7*) as well as all other testa thickness QTLs in case these were also affecting permeability, even though QTLs for permeability were not detected in the *TT4*, *TT5* or *TT7* regions. Genotyping a marker for *Perm7* on already existing material identified two such suitable populations (TT6-a6M and GTY-a5B). Because of a strategic recombination event in both populations permeability could be independently investigated against testa thickness and testa roughness (*TT6a/GTY*) and testa thickness (*TT6b*) QTLs, as shown in Tables 6.3 and 6.4.

As shown in Figure 6.15, neither the *TT6a/GTY* (GTY-a5B) nor the *TT6b* (TT6-a6M) region had a significant effect on permeability. For *TT6a/GTY* region the average time to imbibe for smooth seeded lines (domesticated allele) was calculated at 19.3 days (1.15 days(log)+1). This was 20.46% and 14.26% earlier than homozygous wild and heterozygous plants which imbibed on average in 23.9 and 22.5 days respectively (1.2 and 1.3 days(log)+1). Nevertheless, using a Tukeys pairwise analysis this variation between plants carrying wild and domesticated alleles was calculated as not significant. This indicates that testa roughness had no effect to permeability. As shown in Figure 6.15b, this was further supported when analysing the distribution of plants carrying different *AGO1* alleles in relation to permeability which showed major overlap, implying no correlation.

For the *TT6b* region the average days to imbibe for plants carrying the domesticated allele was 0.82 days (0.8 days(log)+1). This was on average 4.7% and 3.7% earlier than plants carrying wild and heterozygous alleles, which imbibed on average 0.9 and 0.8 days respectively (0.8 and 0.8 days(log)+1). Again, using the Tukeys pairwise analysis the increased permeability in the domesticated genotype could not be validated using the Tukeys pairwise analysis. This was further supported when analysing the distribution of *COP13* alleles in relation to permeability, which as shown in Figure 6.15d, exhibited major overlap, implying no correlation. Collectively these findings confirm the TT6 region which included testa thickness and testa roughness QTLs had no effect to permeability.

It is interesting to note *TT6b* (TT6-a6M) population being significantly more permeable than *TT6a* (GTY-a5B), despite having the same genetic background for permeability QTLs (*Perm2* = wild and *Perm7* = domesticated). This cannot be explained by variation in testa thickness as both populations are relatively similar, with plants carrying the wild allele averaging at 124µm and 127µm (2µm) and plants carrying domesticated allele averaging at 117µm and 119µm (2µm) for the *TT6a* and *TT6b* populations respectively, giving further support testa thickness is being controlled independently to permeability. While no clear explanation can be given, it is possible that this may have resulted from accumulative effect of additional genetic regions not detected in the original QTL analysis in Chapter 4 (PEV = 41.3%) or alternatively markers used to define QTL boundary were not sufficient. Because of the near free germinating habit of the *TT6b* population, which mirrored phenotype of the domesticated *Perm2* phenotype, it is possible the causative gene was closely linked but independent to the suspected loss of pigmentation (Mendel *A* gene), a concept highlighted in Chapter 5. To explore this possibility, the population was screened for closely linked flanking markers (*LF*, *NAD* and *ARGJ*), which were found to have the wild genotype consistent with the Mendel *A* gene (data not shown). This supports both populations having the same *Perm2* genotype and therefore was unlikely to be responsible for the highly permeable phenotype of the *TT6b* population.

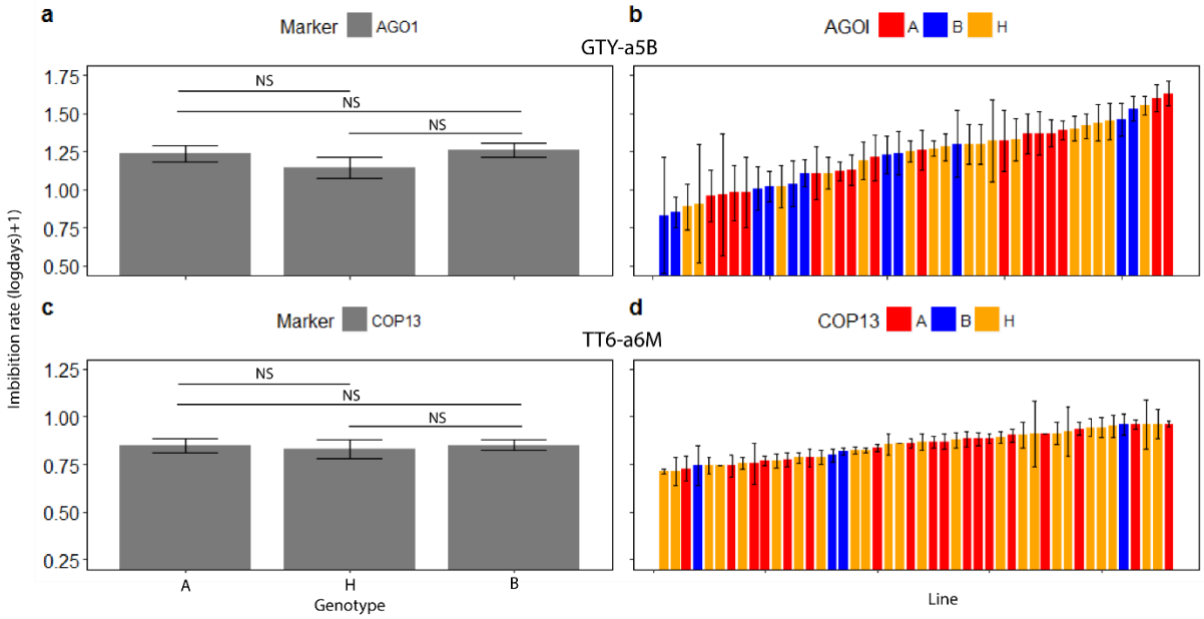


*Figure 6. 15: Effects of permeability on advanced segregating populations GTY-a6M (a-b). TT6-a6M (b-c) Figures a and c show average permeability across population for each marker while Figures b and d show distribution of representative marker based on permeability across population. Error bars are calculated using standard error. Significant differences in permeability between genotypes was calculated using Tukeys pairwise (NS >0.05, * P < 0.05, ** P<0.01, ***P<0.001)*

6.4 Discussion

The testa structure has been implicated as a major component in seed dormancy (Liu et al., 2007, Hradilová et al., 2017, Murphy and Fuller, 2017, Weeden, 2018). According to Plitmann and Kislev (1989) the domestication process has resulted in an average thinning of the testa by 10-25%, while in legumes a 40-45% reduction was reported in *L. culinaris* and over 50% in *C. arietinum* and *P. sativum L*. This was consistent with our findings, which observed a reduction of 55.4% between parental lines, with wild line exhibiting a thickness of 171±3 µm and domesticated 76±2 µm. This concurrent reduction in testa thickness and increased permeability strongly implicates its association to hardseededness, (Lush and Evans, 1980, Debeaujon et al., 2000, Miao et al., 2001, Smykal et al., 2014, Hradilová et al., 2017). This is consistent with the significant reduction in domesticated legume species with their wild progenitor recognised for their discernible thick testa and dominant physical seed dormancy properties. This was supported by findings in Chapter 5 which found a strong co-location of major permeability and major testa thickness QTLs on LGII. It was surprising therefore to find the additional testa thickness QTLs identified in the RIL population did not appear to influence permeability. It is not clear whether this was caused by limitations in the QTL analysis detectability, or whether this indicates that testa thickness plays a less fundamental role in modulating permeability than previously thought. This study explored these concepts in relation to *TT6* on LGVI. This region, referred to as Clst6, was of interest due to its known association with testa roughness, a trait which has often been considered a fundamental feature of domestication in pea due to the ubiquitous

expression of smooth seededness across the domesticated germplasm (Zaytseva et al., 2017). While its functional role has not been explicitly explored, potential links to dormancy have been (Lush and Evans, 1980, Miao et al., 2001) because of its rapid evolution during the pristine or early domestication period.

6.4.1 Fine mapping GTY

Testa roughness has been recognised as a monogenic trait (termed *Gritty*) and has been mapped, albeit in relatively low resolution to LGVI (Wolko and Weeden, 1990, Weeden et al., 1998, Weeden et al., 2002, Weeden, 2007). This study has attempted to refine the map position *GTY,* in part to enable a clearer analysis of its potential effect on properties potentially related to seed dormancy. Previous QTL analysis in Chapter 4 based on a 5-step categorisation of testa roughness confirmed a major QTL on LGVI but also identified a second novel QTL on LGVII near testa thickness QTL, *TT7*. The co-location of QTL for testa thickness and testa roughness in two distinct genomic regions suggests that the two traits might be pleiotropic effects of the same underlying gene, or simply that increased seed roughness influences overall testa thickness as measured by a micrometer expression based on its proximity to both testa roughness QTLs. This is supported by a significant positive correlation between increased testa roughness with increased thickness in the *TT2*, *TT4* and *TT5* regions as shown in supplementary Figure 6.4.

Attempts to isolate this QTL by mapping as a qualitative trait has determined a conservative position between markers *FULa* and *CABB*, an interval totalling 13 recombinations. This region which encompassed *AGO1,* is the most comprehensive mapping of this locus and has been refined to a 6.1cM region according to the Tayeh et al. (2015) linkage map, which relation to *M. truncatula* reference genome comprised of 48 genes (see Supplementary Figure 6.3). Further refinement of this region was precluded by disagreement in mapping position within different populations, findings also reported by DeMason and Weeden (2006) where *GTY* was mapped either of *AGO1*. In this study, *GTY* was most consistently positioned between markers *AGO1* and *CABB* (4.8cM) where it had been independently mapped 13 times, compared to 3 times between *AGO1* and *CABB* (1.3cM).

The nature of this mapping discrepancy is unknown; however, it was unlikely to be caused by genotyping or phenotyping errors as these were validated by growing subsequent populations, with six populations confirming its position between *AGO1* and *CABB* and one population confirming its position between *AGO1* and *FULa* (Figure 6.9a). One possible explanation for its ambiguous map position could be genomic structural changes between the parental populations, causing alternative mapping within different populations. Genomic translocations have been reported between wild and domesticated pea lines (Ben Ze'ev and Zohary, 1973) although it is unclear exactly where these events

occur. It is known that the region where *GTY* has been mapped interval is generally more prone to genomic rearrangements on an evolutionary timescale, with poor synteny reported in other legume species (Kaló et al., 2004, Bordat et al., 2011, Tayeh et al., 2015). This was consistent with this study which showed the *GTY* locus spanned across *M. truncatula* chromosome 2 and 6 with insertions from additional chromosomes and major inversions and genomic rearrangements between this junction. Poor synteny was also observed with its most closely related sequenced genome *L. culinaris*, which if the recently developed genome can be trusted showed independent genomic rearrangements to *M. truncatula*. This further illustrating its susceptibility to genomic rearrangements. In addition, markers in this region showed significant segregation distortion, which may also indicate genomic structural variations within this region.

6.4.2 Validation of testa thickness

Using phenotypic data collected in Chapter 4, an in-depth inspection of the *TT6* region identified two potentially independent testa thickness QTLs. These were referred to as *TT6a* which was defined by markers *FULa* and *CABB* and *TT6b* which was defined by markers *APRL* and *COP13*, although it is possible that this QTL may extend beyond *COP13*. Using advanced generation segregating populations to isolate these regions, both QTLs were validated. The *TT6a* and *TT6b* QTLs both had the same allelic direction, with domesticated alleles at either QTL conferring a 5-8% reduction in testa thickness relative to the wild allele.

TT6a could be definitively mapped between *PhyB* and *CABB* which based on Tayeh et al. (2015) linkage map was measured at 20.8cM. Although further refinement of this interval was precluded by limited recombination events, it seemed most strongly associated with the *AGO1-CABB* region. This coincided with the strongest consensus map position of *GTY*, which implies that these two traits may be manifestations of the same underlying gene. This idea was further supported by the observation that with the testa roughness expression was to some extent associated with other testa thickness QTLs (particularly *TT7*), as previously discussed. An association between testa thickness and testa roughness has been proposed for several legume species (Plitmann and Kislev, 1989), with Miao et al. (2001) reporting roughseeded lines generally had a thicker palisade cell layer than smoothseeded lines in lupin, which may exacerbate any surface texture effects, however there had been no genetic evidence for this association. This has been the first study to isolate *GTY* within a well-defined interval and compare its effect on testa thickness in relatively near-isogenic material. It will be interesting in future to further refine the position of *GTY* and test this association even more rigorously in larger populations.

For *TT6b*, *MTIC153* was identified as a lower QTL boundary which distinguished this QTL from *TT6a*. This definitively showed that the second testa thickness QTL was independent to *GTY* with all lines exhibiting a smooth seeded phenotype. Although an upper QTL boundary could not be defined, testa thickness most strongly associated with the *APRL* and *COP13* marker, suggesting a map position near or above these marker. Once again, larger populations and additional markers will be needed in future to improve this position.

6.4.3 Effect of testa roughness and thickness to permeability

Hardseededness has been linked to both testa roughness in pea (Cechová et al., 2017) and soybean (Otobe et al., 2015), and increased testa thickness across most species (Sefa-Dedeh and Stanley, 1979, Debeaujon et al., 2000, Weeden, 2007, Hradilová et al., 2017). However allelic variation at testa roughness and testa thickness QTLs on LGVI examined in this study showed no significant effect on permeability, consistent with findings from the original QTL analysis in Chapter 4. This lack of association is also consistent with findings in cowpea where smooth seeded cowpea lines actually exhibited a thicker and less permeable testa (Lush and Evans, 1980) compared to rough testas. This is the first study that has isolated the *GTY* region and directly tested its effect on permeability within a background fixed for testa thickness and permeability QTLs. The non-significant effect of testa roughness to permeability was a surprising result. This was because smooth seededness is universally expressed in both independent domesticated pea lineages *P. sativum var sativum* and *P. sativum var abyssinicum* (Zohary and Hopf, 1973, Zohary et al., 2012, Zaytseva et al., 2017), therefore implying a fundamental role in domestication which was assumed to be related to physical seed dormancy properties. One possible selective advantage for smooth seededness may have been for improved palatability, however this does not easily explain why it has undergone such a severe selective sweep in both domesticated pea lineages. These findings do not disprove the possibility of a link between permeability and testa thickness, because as discussed in chapter 5, both traits demonstrated a strong correlation on LGII. Overall these results suggest a complex relationship between testa thickness, orientation and permeability, with some genes potentially contributing to only one trait and others affecting more than one. A better understanding of the physical and anatomical basis for these effects may be achieved in future by the eventual cloning and functional analysis of the causal genes.

In combination with Chapter 4, this study has demonstrated that approximately 50-60% of the variation in testa thickness between the representative wild and domesticated pea lines could be attributed to six QTLs on LGII, IV, V VI and VII. In each case the domesticated allele was associated with a reduction in thickness, suggesting incremental transformation which likely occurred over a protracted period. This coincides with previous findings in pea where numerous testa thickness loci

were identified (Weeden, 2007) as well as other legume species such as hyacinth bean (*Lablab purpureus*) (Maass, 2006) and horsegram (*Macrotyloma uniflorum*) (Murphy and Fuller, 2017). The polygenic control of testa thickness has been argued to indicate dormancy loss in legumes was a protracted rather than a rapid process (Murphy and Fuller, 2017). However, the fact that that this thesis has revealed the potential for testa thickness to be controlled independently of permeability leaves open the possibility that that loss of seed dormancy was a rapid evolutionary process (Ladizinsky, 1987, Abbo et al., 2011, Abbo et al., 2014).

Based on the collective findings in this study and in chapter 5 it seems plausible that the chemical composition of the testa has a more significant role in regulating permeability than testa thickness or testa roughness. Permeability has been shown to be strongly associated with the chemical composition of the cuticle and suberin composition, which normally provides an impermeable barrier (Ma et al., 2004b, Guo et al., 2015). Cutin and suberin induced biosynthesis mutants exhibit increased level cracking within the palisade cell layer that leads to increased permeability (Chai et al., 2016, Lashbrooke et al., 2016, Gou et al., 2017) irrespective of testa thickness (De Giorgi et al., 2015). Vu et al. (2014) reported an increase level of microcracking within domesticated soybean compared to its wild counterparts, concluding testa cracking may have been consciously or unconsciously selected during domestication. This has been supported by characterisation and isolation of domestication related QTLs with induced mutant lines resulting in perturbations to the cuticle formation and increased microcracking, which had no effect on testa thickness (Jang et al., 2015, Sun et al., 2015).

6.5 Conclusions

In pea, the reduction in testa thickness has been regarded as central to its domestication, based on the wild relative exhibiting a thick dormant testa (Ojomo, 1972, Murphy and Fuller, 2017, Fuller and Allaby, 2018). In contrast to these common perspectives, chapter 4 indicated testa thickness may not necessarily be tightly associated with dormancy. This was surprising, particularly for *TT6* which had been mapped near the testa structural locus *GTY*, which appears largely responsible the smooth seeded phenotype present in virtually all domesticated pea lines. During the fine mapping of this region two significant testa thickness QTLs were identified, indicating that a total of at least six QTLs controlling testa thickness within the NGB5839xJI1794 population. Analysing the effect of these QTLs in a near isogenic population with regards to permeability and testa thickness QTLs, no significant effect on imbibition rate was found. This further supports and strengthens the conclusions from chapter 4, that both testa thickness and testa roughness have a less fundamental role in controlling seed dormancy than previously considered. We therefore conclude that its chemical composition is

likely to have a more prominent role in regulating seed dormancy (Souza and Marcos-Filho, 2001), and this will be an interesting topic for future analysis.

This study was unable to identify the causative genes for testa roughness and/or testa thickness in the *TT6a* and *TT6b* regions. This was partly because mapping resolution was by relatively small populations sizes, and because poor synteny between reference genomes prevented an effective candidate gene approach. Nevertheless, this study has provided the clearest and most rigorous mapping of *GTY* to date and established genetic material that will be useful for the further refinement of this position, and potential identification of the causal genomic change. A small number of ambiguous, mutually contradictory mapping results were identified and in at least one case, seemed to be robust on further progeny testing. The explanation for this is unclear but one possibility is that it might relate to genomic structural variations between wild and domesticated parent. This was based on its known susceptibility for genomic rearrangements within this region and observed segregation distortion within different populations. This was concept was not further investigated in this study but would be an interesting topic for future research.

Chapter 7: Genetic dissection of flowering time QTLs

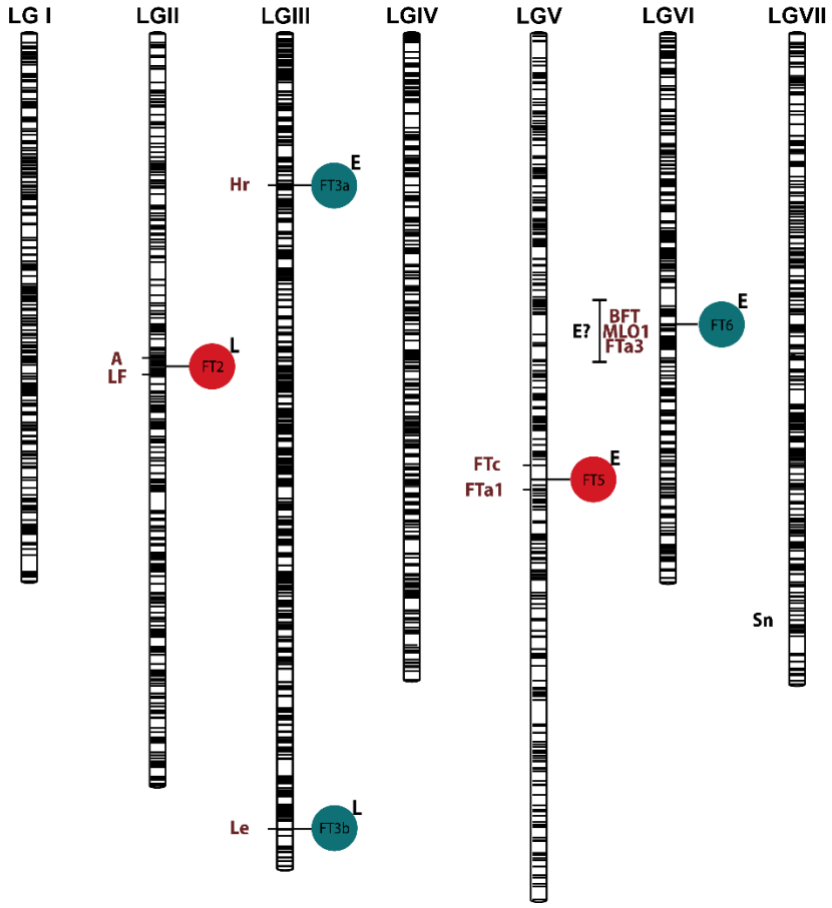
7.1 Introduction

In many crop species, flowering time is a key trait that has been modified through domestication and expansion into new environments. Most crops exhibit earlier flowering to their wild progenitor, this often reflects the relaxation of inhibiting mechanisms related to photoperiod and vernalisation responses, which has permitted expansion beyond the normal eco-geographical range of the wild (Weller et al., 2012, Li et al., 2014a, Nakamichi, 2014). In pea, the wild lineage has a broad latitudinal range extending from Iran and Turkmenistan through Anterior Asia, northern Africa and southern Europe (Smýkal et al., 2011), but is restricted latitudinally due to photoperiod and frost tolerance requirements; while these restriction have been overcome in the domesticated varieties, which have spread both north and south of the wild range. Understanding these underlying genetic components is of interest as this can be used to inform the expansion of the domestication lineage. Furthermore, this study can be utilised in crop breeding programs to tailor lines to flower at specific times to maximise productivity at a site-specific level.

Previous analysis have identified over 20 loci related to flowering in pea (Weller and Ortega, 2015), with natural variation at four loci shown to occur within the domesticated germplasm (Murfet, 1971, Murfet, 1973, Weller et al., 1997, Weeden, 2007, Weller, 2007). These loci include *LATE FLOWERING* (*Lf*) on LGII, *HIGH RESPONSE* (*Hr*) on LGIII, *EARLY* (*E*) on LGVI and *STERILE NODES* (*Sn*) on LGVII. Dominant *Lf*, *Hr* and *Sn* alleles are known to inhibit flowering, while the dominant allele at the *E* locus confers an earlier flowering phenotype in most backgrounds, but can be subject to complex interactions with the other three loci (Weller and Ortega, 2015). The *Hr* and *Sn* loci have subsequently been identified as the circadian clock genes *EARLY FLOWERING 3a* (*ELF3a*) and *LUX ARRHYTHMO (LUX),* respectively, with the recessive *hr*/*elf3a* lines showing partial photoperiod insensitivity (Weller et al., 2012), while *sn/lux* mutant lines exhibit complete photoperiod insensitivity (Liew et al., 2014). The *Lf* locus has been identified as *TFL1c*, a homolog of the arabidopsis *TERMINAL FLOWER 1* gene that has shown to delay flowering in proportion to its expression level in both long and short-days (Foucher et al., 2003). Numerous naturally occurring alleles have been identified which have subsequently been characterised into four allelic classes based on their phenotypic effects and dominance relationships, with dominance and ability to deter flowering in descending order shown as Lf^{d}>*Lf*>*lf*>lf^{a} (Murfet, 1975). The *E* locus has not been characterised at the molecular level but has been shown to promote flowering in both *lf* or *lf-a* background without affecting photoperiod sensitivity (Weller et al., 1997). By studying the distribution of the flowering time mutants across the domesticated germplasm, the early *hr* allele was found to originate from the Mediterranean Europe,

permitted its expansion into northern Europe (Weller et al., 2012). The mutant *sn* was found within a subset of the lines carrying the *hr* allele (Liew et al., 2014) therefore implying a later adoption. While the distribution of *LF* and *E* variants have not been studied, the fact that natural variation at these loci is widespread among cultivated lines suggest that these were non-critical domestication traits.

Chapter 4 described the identification of five flowering time QTLs in the JI1794 x NGB5839 RIL population grown in LD conditions; *DTF2*, *DTF3a*, *DTF3b*, *DTF5* and *DTF6*. A summary of their map positions is shown in Figure 7.1. A previous study using the F_2 generation of the same cross grown under short-day conditions identified QTLs in the position of *DTF3a* and *DTF6* (*QTL3* and *QTL6*) demonstrated that *QTL3* was the *Hr*/*ELF3* gene (Weller et al., 2012). *QTL6* was mapped near to the *E* locus and is currently under investigation in a separate study. The *DTF3b* locus mapped very close to Mendel *Le* gene, which encodes a key biosynthetic protein for production of gibberellin (Lester et al., 1997), a key modulator of plant growth and development. NGB5839 is a dwarf line known to carry an induced mutant allele (*le-3*) (Lester et al., 1999), and this is reflected in the detection of QTL for multiple traits in the region of *Le* in Chapter 4 of this study. This includes earlier flowering time (DTF) with mutant *le* allele, although no significant effect was found to node of flowering (*FN*), findings consistent with Murfet and Reid (1987). *DTF2* mapped close to the known location of *Lf,* and the fact that naturally occurring variants at *Lf* are known to exist implicates this as a strong candidate. However, as the domesticated allele was associated with later flowering (Chapter 4), this implies either; a potential gain of function mutant, which has previously not been reported for *Lf* (Taylor and Murfet, 1993, Foucher et al., 2003, Weller, 2007), or an alternative but closely linked causative gene. *DTF5* mapped near to the known *FT* homolog cluster *FTa1-FTa2-FTc,* which is highly conserved throughout the galegoid legumes (Hecht et al., 2011, Weller and Ortega, 2015). Induced mutants for *FTa1* (*PsGIGAS*) have shown that this gene plays an important role in modulating flowering in both pea and Medicago (Hecht et al., 2011, Laurie et al., 2011), and both *FTa1* and *FTc* are strong promoters of flowering when expressed in transgenic arabidopsis. The domesticated allele at *DTF5* was associated with earlier flowering time, suggesting that if this locus corresponded to a gene in the *FT* cluster, then the domesticated form might represent a gain of function mutation. Interestingly there have been no naturally occurring *gigas* alleles described, or any other natural variation in this region known to affect flowering time.



QTL	LG	Days to flower (DTF)			Flowering node (FN)			Associated early flowering genotype
		Position	PEV (%)	Closest anchor marker	Position	PEV (%)	Closest anchor marker	
DTF2	II	108.928	6.2	Between A (+9.301cM) and LF (3.072cM)	112.305	9.4	LF (-0.305cM)	A
DTF3a	III	51.618	12.6	HR (-0.06cM)	51.618	9.3	HR (-0.03cM)	B
DTF3b	III	267.94	6.8	LE (-3.609cM)	-	-	-	A
DTF5	V	140.519	5.5	FTc (0.00cM)	148.533	6.3	FTa1 (-4.707cM)	B
DTF6	VI	98.022	19.1	Between BFT/MTIC153/NT6083 (+4.651cM) and MLO1 (-0.04cM)	98.564	31	FTa3 (0.00cM)	B

Figure 7. 1: Schematic representation of mapped flowering time QTLs identified in RILs. Red circles signify QTLs of interest, domesticated allelic effect shown as early (E) or late (L) flowering relative to wild. Known marker positions shown in brown text while black text unknown markers. Data summarised in Table.

This chapter intends to clarify the location and effect of *DTF2* and *DTF5*, as well as explore their molecular identity. These loci were not identified in a previous study of the same cross under SD (Weller et al., 2012) and so validating their effects in advanced generation segregating populations will be important.

7.2 Methods

7.2.1 Phenotyping and genotyping

Flowering time was measured by counting number of nodes until the first flower on the main stem. DNA was extracted using standard protocols described in Chapter 2. Genotyping was conducted on *DTF2* and *DTF5* region using standard genotyping protocols described in Chapter 2. Details of markers used are detailed in Table 7.1.

Table 7. 1: Details of target regions and HRM markers used.

QTL	Marker	Primer sequence	Genotyping method
DTF2	*LF*	F: GGTCCCTCTTTACCCTGGTATT R: TGATCTGCAGGAAAACAATAAA	HRM
DTF5	*FTa1*	F: GGACGTGAGCAAAACGACAT R: TTGAGTAGTACCAGCACACACT	HRM
	FTc	F: CATTGGGATGTTAAAATGGTG R: TGGGAAAGAGTTGCAAGATG	HRM

7.2.2 RNA extraction and qPCR

Plant material was grown in normal long-day conditions in a controlled glasshouse environment. Leaf and/or apex tissue samples were collected when leaves had fully expanded at node 9 and at node 14. For each sample, tissue was taken from two individual plants collected from the two near isogenic populations.

RNA extraction and cDNA synthesis were performed using standard protocols described in Chapter 2. Primer sequences used in this chapter are indicated in Table 7.2.

Table 7. 2: Primer sequences used in expression analysis for target genes

Gene	Primer Sequence
ACTIN	F: GTGTCTGGATTGGAGGATCAATC R: GGCCACGCTCATCATATTCA
LF	F: CAGACATTCCAGGGACAACAG R: AAATAAGCAGCAGCAACAGGG
FTA1	F: GCCCAAGCAACCCTACTTTT R: CCATCCTGGAGCGTAAACCC
FTA2	F: GGAAATGACCCCGTGATCTA R: TGAATCCCTAAGTTGGGTCG
FTC	F: CGACTACCGGGACAGCATTT R: CAGGTGAACCAAGGTTATAAAC
RMK	F: TCATGATATAATCCCTGCACTCC R: TGAACCTTCGTTGCTGAACC

7.3 Results

7.3.1 Development of advanced segregating populations

The approach used to validate flowering time QTLs was like the one used in Chapters 5 and 6 whereby advanced generation segregating populations were developed by preferentially selecting and

advancing lines from the F_2 population. In the F_2 population, representative gene-based markers were available for all target QTLs as detailed in Table 7.3, and were selected based on their proximity to QTL peak markers. Of the five QTLs, only *DTF3b* which likely represents the *Le* gene was not a critical target as this showed no effect on node of flowering. Two representative markers for *DTF2* spanned the peak marker and included the strong candidate gene *LF*. For *DTF3a* and *DTF3b* the likely causal genes (*Hr/ELF3a* and *Le*, respectively) were used as representative markers. For *DTF5* the two candidate genes *FTa1* and *FTc* were used as representative markers, whereas for *DTF6* a single representative marker was used (*MLO1*).

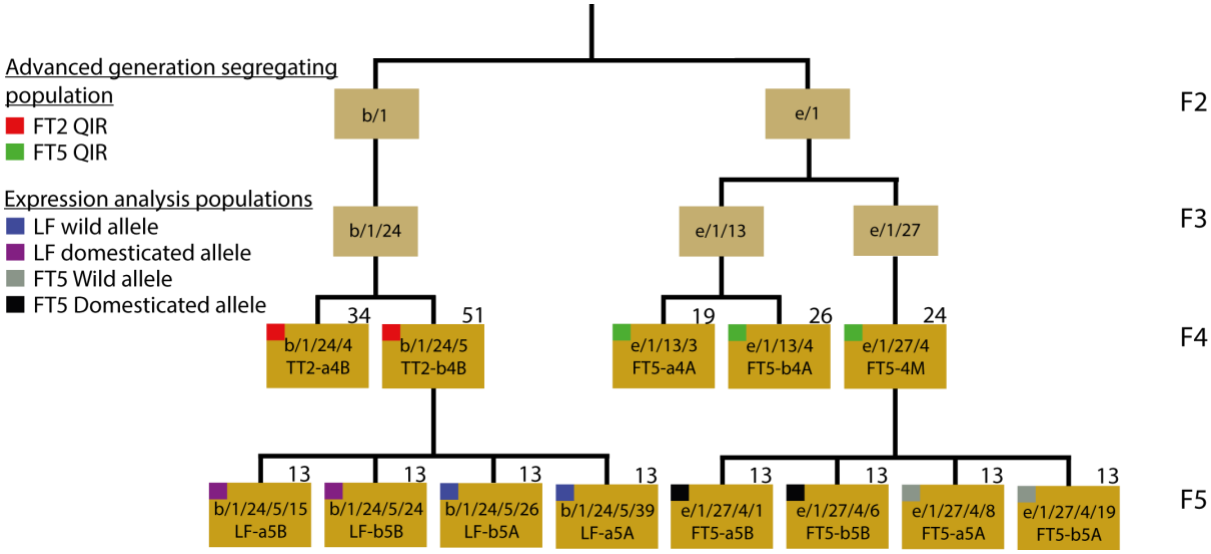
Table 7. 3: Flowering time peak QTL markers identified in RIL population and representative markers positions from F2 population mapped in relation to M. truncatula Mt4.0 v1.0 genome. Marker positions are represented as gene number and the start position of start gene.

	DTF2			***DTF3*a**			***DTF3*b**			***DTF5***			***DTF6***		
QTL	Peak DArT marker		Medicago Position	Peak DArT marker		Medicago Position	Peak DArT marker		Medicago Position	Peak DArT marker		Medicago Position	Peak DArT marker		Medicago Position
	DFT	3540348_1	1g066760 chr1 28.68Mb	DTF	3564019_4	3g104500 chr3 48.14Mb	DTF	4661775_2	2g102740 chr2 44.27Mb	DTF	*FTc*	7g085040 chr7 32.87Mb	DTF	3548711_1	-
	FN	4657639_1	1g060430 Chr1 26.30Mb	FN	3564019_4	3g104500 chr3 48.14Mb				FN	3562879_3	7g083840 chr7 32.28Mb	FN	*FTa3*	6g033040 chr6 10.37Mb
Representative marker	Marker		Position	Marker		Position	Marker		Position	Marker		Position	Marker		Position
	LF		1g60190 chr1 26.20Mb	*Hr*		3g103970 chr3 47.90Mb	*Le*		2g102570 Chr2 44.20Mb	*FTc*		7g085040 chr7 32.87Mb	*MLO*		6g033330 chr6 10.66Mb
	A		1g072320 chr1 32.09Mb							*FTa1*		7g084970 chr7 32.84Mb			

As shown in Figure 7.2, QTL validation was performed in segregating F_4 progeny, while expression analysis compared homozygous F_5 families that differed only at the locus in question.

The *DTF5* populations were derived from two independent F_2 lines e/1/13 and e/1/27. Line e/1/13 was fixed with a wild background for all target QTLs with two lines advanced from F_3 to F_4 to generate populations FT5-a4A and FT5-b4A. The second line (e/1/27) was not fixed for all target QTLs in the F_2 and genotyping in the F_3 was required to develop F_4 population FT5-4M. This was fixed for all target QTLs apart from *FT3b* which as discussed earlier should not affect flowering node results. The *DTF2* advanced generation segregating populations were derived from a single F_2 line b/1/24. These populations had previously been developed in Chapter 5 when isolating permeability QTLs, but incidentally were fixed for all flowering time target QTLs.

The *DTF2* and *DTF5* expression analysis populations were derived from single corresponding F_4 population. This included two lines with a wild genotype and two lines with a domesticated genotype in corresponding QTL regions for each locus.



				QTL and representative markers							
				DTF2		*FT3a*	*FT3b*	*DTF5*		*DTF6*	
Advanced segregating population	Population	Parent line	Generation	*LF*	*A*	*ELF3*	*LE*	*FTa1*	*FTc*	*MLO*	Population Size
DTF2 QTL validation	TT2-a4B	b/1/21/4	F_4	H	D	A	H	A	A	B	34
	TT2-b4B	b/1/24/5	F_4	H	D	A	H	A	A	B	51
DTF5 QTL validation	FT5-a4A	e/1/13/3	F_4	A	D	A	A	H	H	A	19
	FT5-b4A	e/1/13/4	F_4	A	D	A	A	H	H	A	26
	FT5-4M	e/1/27/4	F_4	B	B	A	H	H	H	A	24
DTF2 Expression analysis	LF-a5A	b/1/24/5/39	F_5	A	A	A	A	A	A	B	13
	LF-b5A	b/1/24/5/26	F_5	A	A	A	A	A	A	B	13
	LF-a5B	b/1/24/5/15	F_5	B	B	A	A	A	A	B	13
	LF-b5B	b/1/24/5/24	F_5	B	B	A	A	A	A	B	13
DTF5 Expression analysis	FT5-a5A	e/1/27/4/8	F_5	A	A	A	A	A	A	A	13
	FT5-b5A	e/1/27/4/19	F_5	A	A	A	A	A	A	A	13
	FT5-a5B	e/1/27/4/1	F_5	A	A	A	A	B	B	A	13
	FT5-b5B	e/1/27/4/6	F_5	A	A	A	A	B	B	A	13

Figure 7. 2: Details of all advanced generation population analysed, for QTL validation in the F_4 generation and expression analysis in the F_5. (a) gives a schematic representation of the origin and development of populations, solid lines denote direct link between populations, function of population shown top left corner and population size at top right. (b) Table detailing each advanced.

7.3.2 Flowering time *DTF2*

In Chapter 4, two flowering time measurements, days to flower (*DTF*) and node of flowering (*FN*), identified a QTL on LGII near known flowering time gene *LF*. As shown in Figure 7.3, a detailed inspection of this region illustrates close mapping of *DTF* and *FN* peak markers within a narrow interval of around 3cM, to anchor markers Mendel *A* and *LF*.

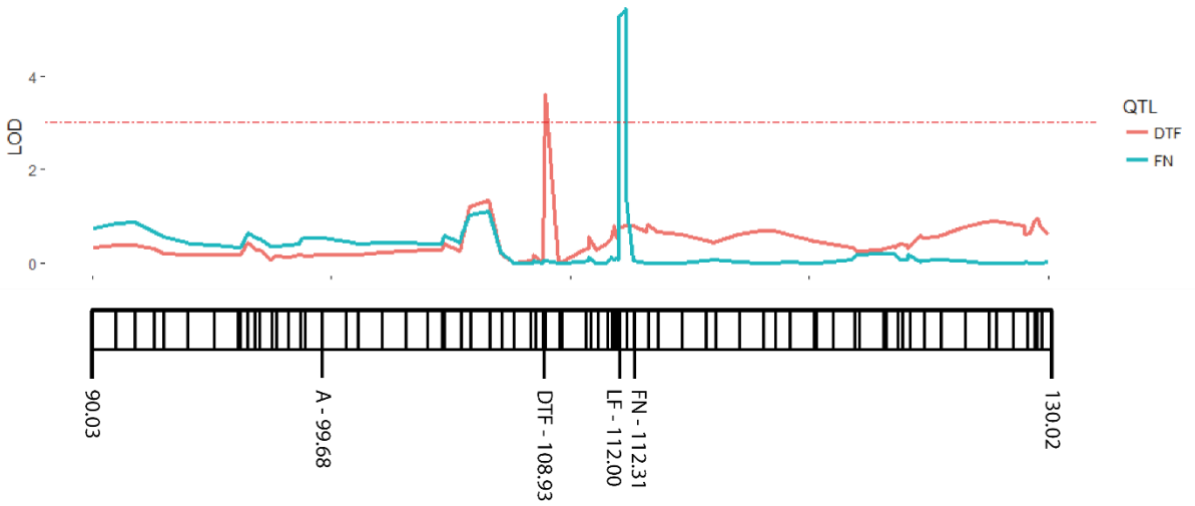


Figure 7. 3: Detailed inspection of the DTF2 region which includes QTL analysis data collected from Chapter 4 for days to flower (DTF) and nodes to flower (FN). Region was compare between 90.026cM to 130.024cM on PsLGII, markers positions across region are shown by black lines, with anchor and peak markers positions given. Dashed red line shows significant LOD 3 threshold

7.3.2.1 Validation of *DTF2*

Figure 7.4a shows that plants carrying the domesticated *DTF2* allele were on average significantly later flowering than plants carrying the wild allele by 3.8 and 3.1 nodes, and later than plants with heterozygous allele by 2.1 and 1.4 nodes in populations TT2ba-4B and TT2b-4B respectively. The phenotype distribution in Figure 7.4b shows clear distinction of the three genotypes in relation to flowering time habit which suggests flowering time is segregating as a single locus. The clear intermediate flowering time habit of plants carrying heterozygous allele indicates incomplete dominance.

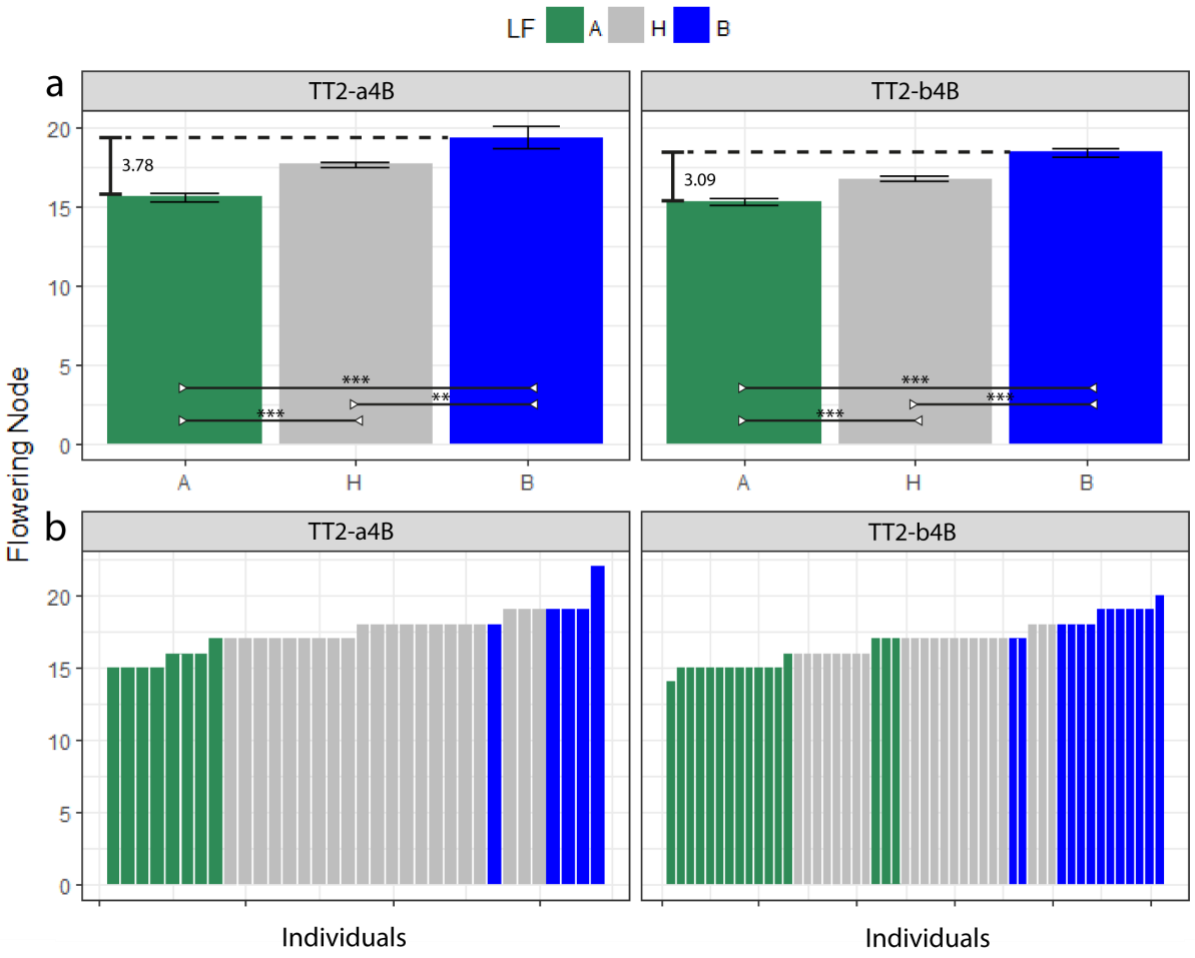


Figure 7. 4: Measuring flowering time across DTF2 advanced generation segregating population (a) average flowering time per genotype, (b) Flowering time per individual. Colours signify genotype Wild = green, Het = Grey, Domesticated = Blue. Dashed line shows difference in average node of flowering between wild and domesticated genotypes. Asterisks indicates significance between genotypes based on Tukeys test: >0.05 (-), <0.05 (*), <0.01 (**), <0.001 (***)

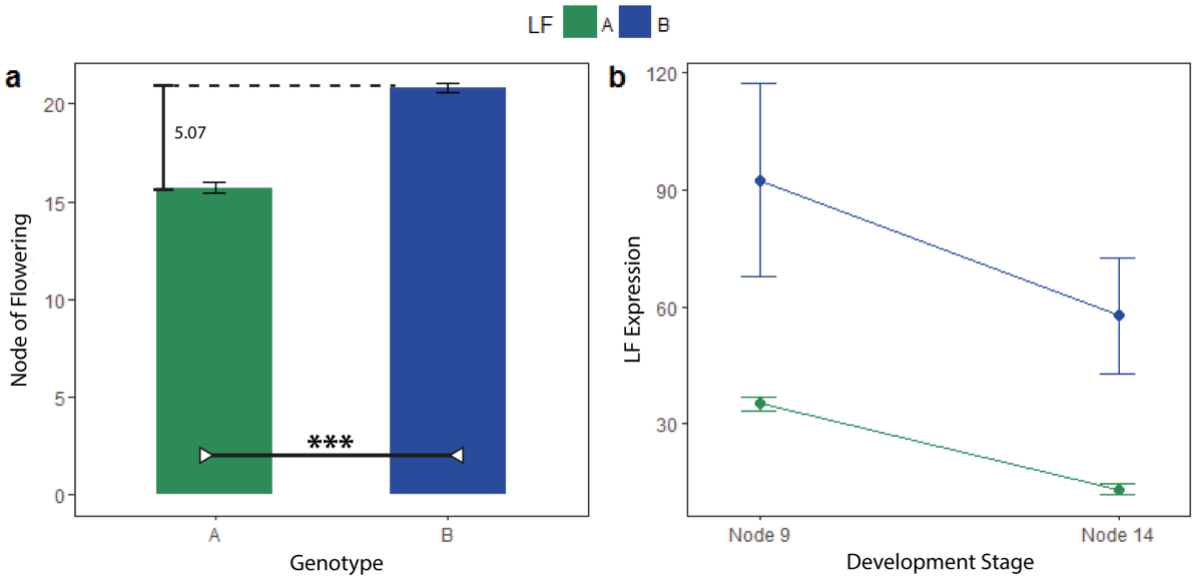
7.3.2.2 Expression analysis of *LF* as candidate for *DTF2*

In view of the known function of *LF* as a floral repressor, its colocation with the QTL peak and its strong association with flowering node in the F_4 population, it was considered a strong candidate for *DTF2*. To investigate its potential role in regulating flowering its expression was compared between lines carrying wild and domesticated alleles as described in Section 7.3.1.

As shown in Figure 7.5a, lines carrying the domesticated *LF* allele flowered significantly later than the lines carrying the wild allele, by approximately 5 nodes. This was consistent with findings from the F_4 generation where the difference between homozygous wild and domesticated genotypes was 3-4 nodes, validating this population having an effect to flowering time.

Expression of *LF* was measured only in the apical tissue, as previous reports had shown its expression was restricted to this region (Foucher et al., 2003). To cover the possibility that expression might change over time, expression was measured at two development stages, at around 9 and 14 leaves

expanded. As shown in Figure 7.5b expression levels were higher at the earlier stage in both genotypes, consistent with its more repressed floral state during early development. Comparing expression between genotypes, *LF* expression was significantly elevated in the domesticated allele at both development stages (Figure 7.5c). This is consistent with the observed delay in the node of flowering, with average node at 15.7±0.3 and 20.8±0.2 for wild and domesticated alleles, respectively.



c

Node 9	Node 14
0.0472	0.0217

Figure 7. 5: Analysis of (a) flowering time and (b) expression plants carrying wild and domesticated DTF2 allele. Flowering time for each genotype was based on an average of 14 plants from two near isogenic populations. Expression analysis was conducted on apical tissue when leaves had fully expanded at nodes 9 and 14. This comprised of three replicates with each replicate including samples collected from both near isogenic line. Dashed line shows difference in average node of flowering between wild and domesticated genotypes. Asterix indicates significance between genotypes based on Tukeys test: >0.05 (-), <0.05 (*), <0.01 (**), <0.001 (***). (c) determines significance in expression between wild and domesticated for different genes/regions at different development stages, using Tukeys pairwise analysis. Significant results (P<0.05) is highlighted in yellow.

7.3.3 Flowering time *DTF5*

In Chapter 4, *DTF* and *FN* identified a QTL on LGV near to flowering time *FT* homolog cluster *FTa1-FTa2-FTc*. Figure 7.6 shows a higher resolution plot of this same data, which suggests a slightly different QTL position for DTF located closer to *FTc*, and FN closer to *FTa1*, although both QTLs were mapped within a narrow interval of around 8cM.

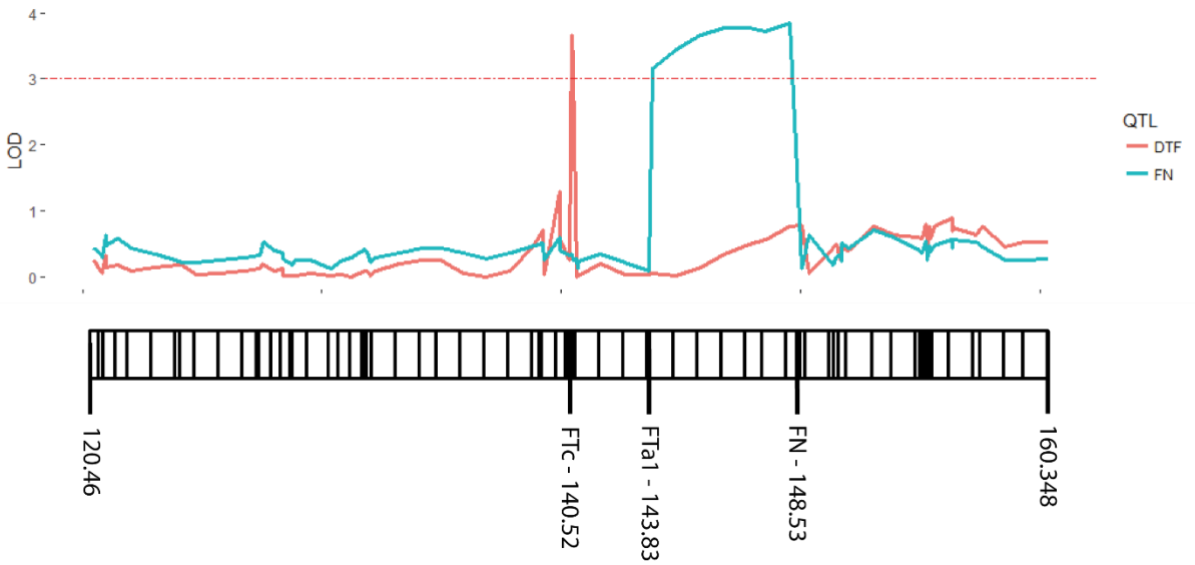


Figure 7. 6: Detailed inspection of the DTF6 region which includes QTL analysis data collected from Chapter 4 for days to flower (DTF) and nodes to flower (FN). Region was compare between 120.456cM to 160.348cM on PsLGVI, marker positions across region shown by black lines, with anchor and peak markers positions given. Dashed red line shows significant LOD 3 threshold

7.3.3.1 Validation of *DTF5*

To validate this QTL, three F_4 advanced generation segregating populations were developed as discussed in section 7.3.1. Populations FT5-a4A and FT5-b4A had the same genetic background in relation to the flowering time QTLs, while FT5-4M differed in the *DTF2* region which carried the domesticated (late) rather than the wild (early) allele (Figure 7.2b). These three populations consisted of 69 individuals in total (19+26+24). Genotyping for *FTa1* and *FTc* revealed no recombinations between markers.

As shown in Figure 7.7a, plants homozygous for the domesticated *DTF5* allele flowered significantly earlier than plants homozygous for the wild allele. This was observed in both genetic backgrounds, with FT5-a4A and FT5-b4A flowering 2.0 and 1.7 nodes earlier while FT5-4M flowering 3.1 nodes earlier. In each family, heterozygous plants had an intermediate flowering node phenotype, that in two of the three families was significantly different from the homozygous genotypes. Figure 7.7b shows that in each family, genotypes can be clearly distinguished in relation to the distribution of flowering node. This indicates that this trait is close to segregating as a Mendelian locus, showing how advanced segregating populations had successfully removed additional major effecting flowering time QTLs.

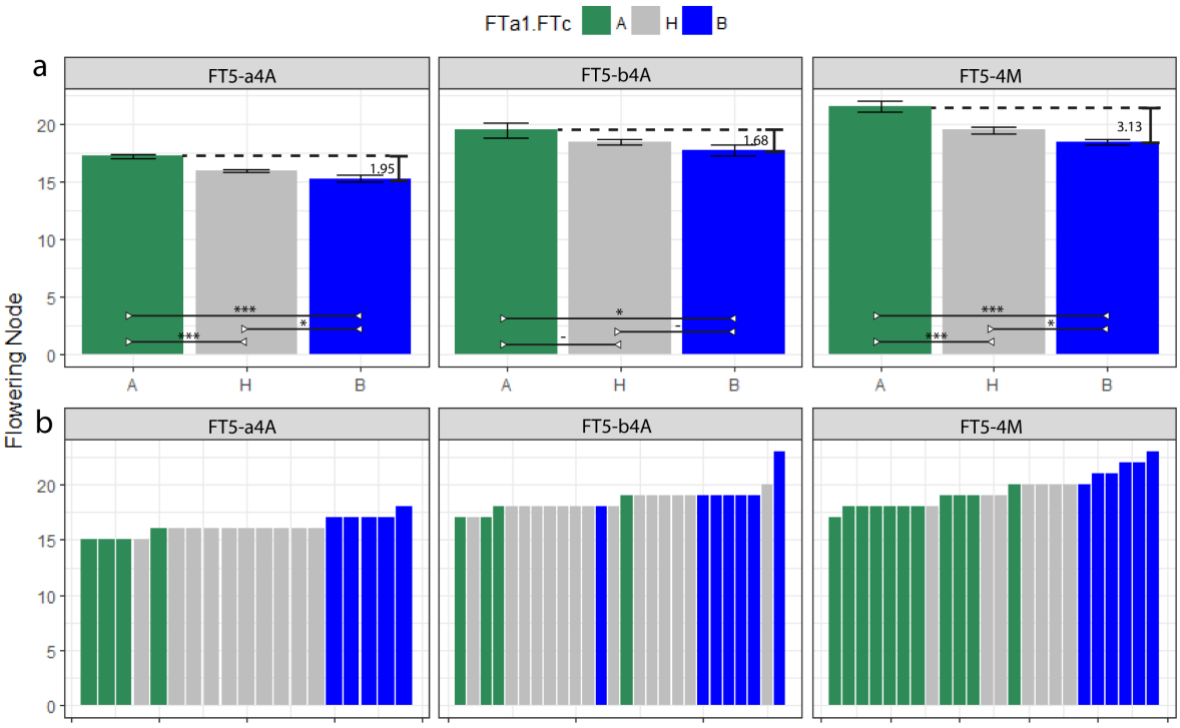


Figure 7. 7: Measuring flowering time across DTF5 advanced generation segregating populations FT5a-4A, FT5b-4A and FT5-4M (a) average flowering time per genotype, (b) Flowering time per individual. Colours signify genotype: Wild = green, Het = Grey, Domesticated = Blue. Dashed line shows difference in average node of flowering between wild and domesticated genotypes. Asterisks indicates significance between genotypes based on Tukeys test: >0.05 (-), <0.05 (*), <0.01 (**), <0.001 (***)

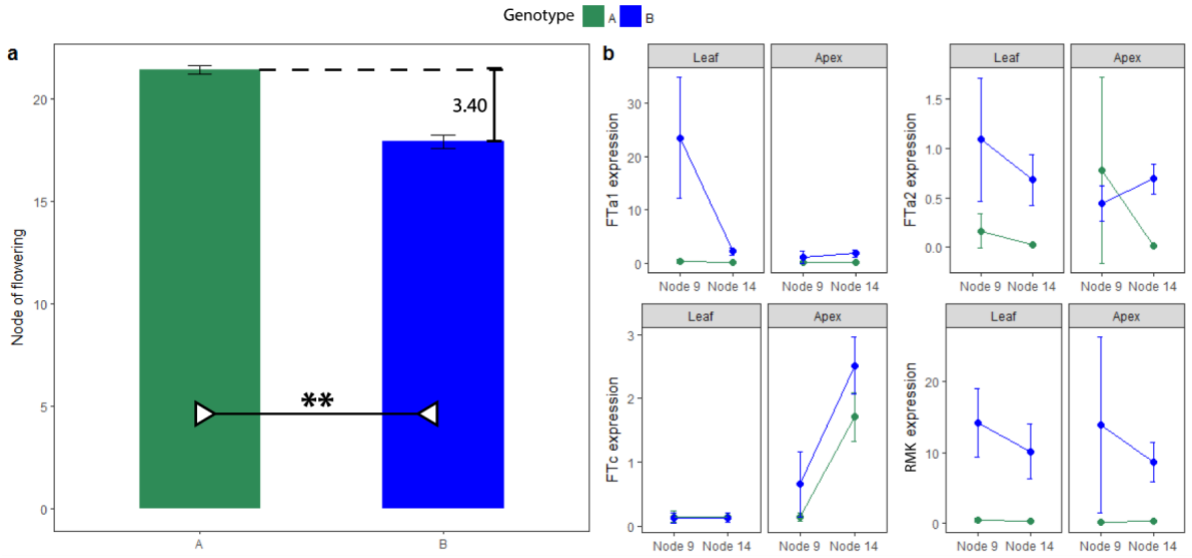
7.3.3.2 Expression analysis of *FTa1/FTc* cluster as candidates for *DTF5*

In consideration of known *FT* function and the strong association of *FTa1* - *FTc* segregation to flowering time observed in the F_4 population, the *FT* cluster constitutes as a strong candidate for *DTF5*. As previously mentioned, no recombination events were identified between *FTa1* and *FTc* markers due to their tight linkage, which in relation to the *M. truncatula* genome extends 0.03Mb region (32.84Mb to 32.87Mb). Consequently, all genes within this region must be considered candidates.

To investigate their potential involvement in the *DTF5* QTL the expression of *FTa1*, *FTa2* and *FTc* was compared between lines carrying wild and domesticated *DTF5* alleles. In addition, an expressed intergenic region between *FTa1* and *FTa2* (referred to as *RMK*) was also examined in view of growing evidence this region regulates *FTa1* (Mauren et al., 2013, Yeoh et al., 2013, Rajandran, 2016, Ortega, 2018). As described in section 7.3.1 four advanced segregating populations were developed. These were near isogenic lines based on flowering time QTLs with two populations carrying the wild (FT5-a5A and FT5-b5A) and two populations carrying the domesticated allele (FT5-a5B and FT5-b5B), (Figure 7.2b). Two near isogenic populations were used in the analysis to reduce effect of additional QTLs that may influence results as mentioned in Section 7.3.2.2. Expression was measured in leaf and

apex tissue at two different developmental stages (leaves fully expanded at node 9 and node 14) to investigate the effect and changes of expression over time and space as previously described.

As shown in Figure 7.8a, phenotyping results were consistent with the F_4 population, showing plants homozygous for the domesticated allele flowered significantly earlier than wild homozygotes, therefore confirming flowering time effect. This was also consistent with its genetic background which flowered 3.4 nodes earlier in the F_5 compared to 3.1 nodes earlier in its F_4 progenitor. Figure 7.8b shows that at both development stages the expression for all three *FT* genes (except *FTa2* in the apex at node 14) and the *RMK* intergenic region were higher in plants carrying the domesticated allele, which is consistent with their earlier flowering phenotype. However, as shown in Figure 7.8c, differences for *FTc* were not significant, whereas *FTa1*, *FTa2* and *RMK* showed a significant difference in both tissues at the node 14 timepoint. At the earlier (node 9) timepoint, the only significant difference was seen for *RMK* expression in leaf tissue.



c

	Leaf		Apex	
	Node 9	Node 14	Node 9	Node 14
FTa1	0.0677	0.0351	0.4094	0.0351
FTa2	0.1527	0.0057	0.6914	0.0057
FTc	0.8285	0.1698	0.2892	0.1698
RMK	0.0248	0.0217	0.2493	0.0217

Figure 7. 8: Analysis of (a) flowering time and (b) expression plants carrying wild and domesticated DTF5 allele. Flowering time for each genotype was based on an average of 14 plants from two near isogenic populations. Expression analysis was conducted on leaf and apex tissue at nodes 9 and 14. This comprised of three replicates with each replicate including samples collected from both near isogenic line. Dashed line shows difference in average node of flowering between wild and domesticated genotypes. Asterix indicates significance between genotypes based on Tukeys test: >0.05 (-), <0.05 (*), <0.01 (**), <0.001 (***). (c) determines significance in expression between wild and domesticated for different genes/regions at different development stages, using Tukeys pairwise analysis. Significant results (P<0.05) is highlighted in yellow.

7.4 Discussion

The timing of flowering is fundamental to a plants reproductive success, and in most temperate adapted species seasonal environmental cues integrated through photoperiod and vernalisation pathways. Domesticated crops tend to exhibit an earlier flowering habit compared to their wild counterpart which is driven by the relaxation in the floral supressing mechanisms. This results in the shortening of the growth cycle causing a shift from winter to spring cropping. This permits an expansion in their range beyond regions normally restricted by photoperiod and vernalisation requirements. The circadian clock genes, such as *ELF3* (Weller et al., 2012, Ridge et al., 2017, Wu et al., 2017) and *CO-FT* pathway genes have been common targets for photoperiod insensitivity (Blackman, 2013, Wu et al., 2013, Song et al., 2017, Soyk et al., 2017, Pickersgill, 2018), while *AP1*-like transcription factor and vernalisation regulator gene *VRN1* shown to be key modulators of vernalisation in cereals and *FLOWERING LOCUS C* (*FLC*) and *FRIDGA* (FRI) in Brassicaceae (Lenser and Theißen, 2013).

In pea it was previously determined that the domesticated line NGB5839 flowered substantially earlier than wild line JI1794 in short-day conditions. This difference was controlled by two known genomic regions; one on LGIII reflecting the action of the *Hr/ELF3a* gene, and a second on LGVI (Weller et al., 2012). Chapter 4 of this thesis showed that both loci also affected flowering in LD, and identified three additional QTLs (*DTF2*, *DTF3b* and *DTF5*) that mapped to other locations. The co-location of *FT3b* with Mendel *Le* gene is consistent with previous reports that *le* and other gibberellin-deficient mutants cause a mild delay in flowering time (Murfet and Reid, 1987).

7.4.1 Investigating molecular control of *DTF2*

As previously described in Chapter 4, a flowering QTL was recognised near the top of LGII, associated to later flowering for plants carrying the domesticated genotype. This was referred to as *DTF2* and mapped near anchor marker *LF* which encodes a floral repressor homolog *PsTFL1c* (Foucher et al., 2003). In arabidopsis, *TFL1* has a dual role of regulating flowering time and determinacy, whereas in legumes expansion of *TFL1* gene has resulted in subfunctionalisation, such that in pea *TFL* homologs *DET* (*TFL1a*) primarily affects shoot determinacy while *LF* (*TFL1c*) primarily affects flowering initiation (Murfet, 1975, Weller et al., 1997, Foucher et al., 2003, Weeden, 2007) homologs. *TFL1* homologs have shown to be targeted during domestication in numerous species including wheat *HvCEN* (Comadran et al., 2012) common bean *FIN*/*PvTFL1y* (Gonzalez et al., 2016b, Bhakta et al., 2017), soybean *GmDt1/TFL1* (Tian et al., 2010, Li et al., 2013), and pigeon pea *CcTFL1* (Mir et al., 2014). However, all these examples affect orthologs of a separate pea gene, *DET*/*TFL1a* (Foucher et al., 2003),

with no reports of the targeting of floral regulatory ortholog *LF/TFL1c* gene, apart from in pea (Weeden, 2007).

In the examples given above the domesticated (mutant) allele for *TFL1a* confers earlier flowering and/or determinate growth phenotypes by loss of function. This was in contrast with our previous findings in Chapter 4, which reported later flowering with plants carrying the domesticated allele, therefore implicating a gain of function. In view of this, *LF* was considered a strong candidate for *DTF2*, despite a previous study using the same cross showing no significant effect in short-day conditions (Weller et al., 2012).

Using advanced generation segregating populations *DTF2* was validated in both the F_4 and F_5 generation, and consistent with results from the RIL analysis in Chapter 4, with plants carrying the domesticated allele flowering significantly earlier than wild. The intermediate flowering time of the plants carrying the heterozygous allele is consistent with the previously described *Lf* allele dose effect (Murfet, 1975, Foucher et al., 2003).

In this genetic background, at both development stages, plants carrying the domesticated allele showed significantly higher levels of *LF* expression than those carrying the wild (63% higher at node 9 and 37% at node 14 in relation to the wild. This constituted to a 24.4% (5.04 nodes) delay in flowering time compared to the wild. This is consistent with the role of *LF* as a floral repressor (Foucher et al., 2003, Hecht et al., 2011), which likely competes with *FTa1* for the binding of FD in a 14-3-3 floral promoting complex (Wickland and Hanzawa, 2015, Kaneko-Suzuki et al., 2018). In addition, a reduction in *LF* expression over time was observed in plants carrying both the wild and domesticated allele. This is consistent with the previous observation that *FTa1* supresses *LF* in the apex as the plant prepares to flower (Hecht et al., 2011).

These results imply the domestication process has resulted in a gain-of-function mutation, despite loss of function being most common during domestication (Meyer and Purugganan, 2013). The nature of this mutation was not characterised in this study, however gain-of function would most likely involve mutation within regulatory sequences, either intronic or intergenic. In *LF*, introns have been implicated in regulating expression, with point mutations within intron 1 (haplotype C) and 3 (haplotype E) associated with early flowering (Foucher et al., 2003). There been no reports of gain of function mutation in *Lf*, with previous work by Weeden (2007) providing no information regarding allelic direction of suspected *LF* flowering time gene in primitive x cultivated population (Cultivated × WL808).

In general crop domestication is usually associated with earlier flowering, an adaptation which is also observed in pea. This has shortened the growing season permitting crop expansion to regions previously restricted by photoperiod requirements (Weller et al., 2012). Despite general selective pressures towards earlier flowering lines within cultivated environments, here we report a domesticated allele conferring later flowering. This may have been an adaptation to regions of greater abiotic stress, like for example with sunflower, where later flowering permitted its southerly expansion from central and eastern America to Argentina with longer growth cycle coinciding with higher seasonal rainfalls and higher yields (Blackman et al., 2011). Conversely, it could also be argued to be an adaptation to less stressful conditions, by extending lifecycle this would permit the plant to capitalise on available resources.

7.4.2 Investigating the molecular control of *DTF5*

As previously described in Chapter 4, plants carrying the domesticated *DTF5* genotype flowered earlier. This region was mapped to LGV near a known *FT* homolog cluster *FTa1-FTa2-FTc*. This is a conserved region throughout the galegoid legumes (Hecht et al., 2005) and these genes are collectively strong candidates, as all *FT* genes when expressed in transgenic *arabidopsis* showed floral promoting capabilities to some extent (Hecht et al., 2011). *FTa1* is of particular relevance as this has shown to be a key floral integrator gene in both pea and Medicago (Hecht et al., 2011, Laurie et al., 2011). Major flowering time QTL with similar properties to *DTF5* (semi-dominant inheritance) been found to map to the corresponding region in chickpea (Weller and Ortega, 2015, Ortega, 2018) and lentil (Rajandran, 2016), but currently none in pea (Weller and Ortega, 2015) with Weller et al. (2012) identifying no effect in short-day conditions using the same cross. In an earlier study, Weeden (2007) identified a flowering time QTL on LGV (*QTL-V*) in a wild x cultivated RIL population (JI1974 x Slow) (Weeden, 2007), which may confer to *DTF5*, however this could not be confirmed due to limitations in mapping resolution. This QTL was not found in either of the other wild/primitive/domesticated crosses, indicating this was not a critical domestication trait.

DTF5 was validated in F_4 and F_5 advanced generation, with plants carrying the domesticated allele flowering significantly earlier than plants carrying the wild and heterozygous alleles. This was observed in two different genetic backgrounds varying for *DTF2*, which as previously discussed altered *LF* expression. This implies *Lf* has an additive effect to *DTF5*, with greater expression of *LF* causing later flowering. This would suggest that the effect of the *DTF2* QTL is not dependent on genotype at the *DTF5* QTL, and that the mechanism by which the domestic *LF* allele is misregulated may not involve the previously identified down-regulation of *LF* by *FTa1* (Hecht et al., 2011).

Figure 7.7b showed flowering time corresponded strongly with genotype, therefore indicating the advanced segregating populations had successfully removed additional major effecting flowering time QTLs. Subsequently confidence is established that variation in flowering time primarily results from *DTF5*. The *DTF5* was genotyped using the *FTa1* and *FTc* which are the flanking genes in the *FTa1-FTa2-FTc* cluster, with genotyping results finding no recombination events between these markers. This was not surprising considering their close physical proximity in closely related species including Medicago v4.1 ~33kb (Phytozome) and chickpea ~57kb (Ortega, 2018).

The results from expression analysis suggest that the earliness associated with the domesticated allele could reflect a higher level of *FTa1* expression. These findings are consistent with previous reports of *FTa1*-overexpressing lines resulting in earliness in Medicago (Yeoh et al., 2011, Yeoh et al., 2013) and linkage disequilibrium studies reporting strong selection between winter and spring pea crops in this region (Siol et al., 2017). *FTa1* is an important integrator of both photoperiod and vernalisation pathway in Medicago and is critical for floral initiation in pea under certain LD conditions (Hecht et al., 2011). It is induced in leaves, and the encoded protein is proposed to act as a mobile florigen molecule travelling to the apex via the phloem where it promotes the expression of *FTc* and floral meristem genes (Hecht et al., 2011, Laurie et al., 2011), probably by physical association with the transcription factor VEG2/FD (Sussmilch et al., 2015).

However, not only *FTa1*, but also the adjacent gene *FTa2* and the expressed intergenic sequence *RMK* were observed to show increased expression in plants carrying the domesticated *DTF5* allele. The variation of expression between the two genotypes remains relatively constant over the two development stages for *FTa2* in the leaf and *FTc* in the apex. This is consistent with findings from Hecht et al. (2011) which showed *FTa2* was epistatic to *FTa1* while *FTc* was a downstream target. Therefore, despite the inherent increase of *FTc* and *FTa2* expression for plants carrying the domesticated allele, these are likely the downstream effects of *FTa1* overexpression.

The nature of this mutation was not characterised in this study however, expression profiles from the *FTa1-FTa2* intergenic region provide clues. As previously described, *FTa1* has a central role in integrating photoperiod and vernalisation environmental cues (Hecht et al., 2011, Laurie et al., 2011), however the upstream mechanism which regulate *FTa1* is poorly understood. In arabidopsis vernalisation is regulated by floral repressor *FLC* (Michaels and Amasino, 1999) which is stably down regulated after cold treatment by chromatin silencing mechanism (Hepworth and Dean, 2015, Wu et al., 2016); while in photoperiod pathway, *FT* is regulated by *CONSTANS* (*CO*) which in turn is regulated by the circadian clock (Simon and Coupland, 1996). Nevertheless, these mechanisms are not conserved in legumes, which is demonstrated by absence of *FLC* homologs (Hecht et al., 2005) and

non-functional *CO* and *COL* genes in relation to flowering (Weller et al., 2009, Putterill et al., 2013), therefore implicating different mechanism involved in modulating these pathways (Fudge et al., 2018).

In legumes the photoperiod and vernalisation mechanisms also regulate *FTa1* via independent pathways (Putterill et al., 2013). In Medicago, constitutive expression of *FTa1* shows both vernalisation and photoperiod insensitive early phenology (Laurie et al., 2011), however certain lines have shown vernalisation insensitive but photoperiod responsive habits (Mauren et al., 2013, Yeoh et al., 2013). Also, insertion mutations within the *FTa1-FTa2* intergenic region in Medicago show increased *FTa1* expression, resulting in vernalisation insensitive but photoperiod responsive early flowering phenotypes (Mauren et al., 2013, Yeoh et al., 2013). Rajandran (2016) identified a major deletion in the *FTa1-FTa2* intergenic region (including the expressed non-coding *RMK* region) in lentil that was associated with early flowering in line ILL2601. Also, in both chickpea and lentil expression of the intergenic region was shown to respond to vernalisation and likely involved in the vernalisation response (Ortega, 2018).

Consistent with chickpea and lentil (Rajandran, 2016, Ortega, 2018) the *RMK* expression profile in this study mirrored that of *FTa1*, supporting evidence that transcription of *RMK* is coupled with *FTa1* (Mauren et al., 2013, Yeoh et al., 2013, Ortega, 2018). Plants carrying the domesticated allele showed upregulation of *RMK* in both leaf and apex tissue, while remaining suppressed in wild. In view of these findings and growing evidence for *RMK* exclusive role in modulating the vernalisation response (Ortega, 2018), it is proposed that earliness in domesticated lineage could result from mutation within the *FTa1-FTa2* intergenic region, implicating relaxation of vernalisation responsive mechanism which normally supressed *FTa1* expression in the absence of vernalisation. This is supported by a preliminary vernalisation study showing that in LD and SD conditions the wild line (JI1794) flowered 3.5 and 9.0 nodes earlier when vernalised respectively, compared to 1.1 and 2.0 nodes for the earlier domesticated line (NGB5839) (J. Vander Schoor unpublished data). Alternative explanations that cannot be ruled out at this stage may include polymorphisms in the *FTa1* promoter that relax a repressive mechanism, as proposed for a major flowering time QTL conferring de-repressed *FT* expression in lupin (Nelson et al. 2017). Another could be duplication or otherwise increased copy number of *FTa1*, as described for the *FT* gene, *HvFT1,* in the dominant spring flowering Barley line BGS213 (Nitcher et al., 2013).

Further analysis including sequencing of *FTa1*/*FTa2* intergenic and *FTa1* upstream regions is clearly required in future to distinguish among these possibilities and will be greatly assisted by completion of the pea genome sequence, as very little genomic sequence is currently available.

7.5 Conclusions

Results presented in this chapter validated domestication-related flowering time QTLs: *DTF2* and *DTF5*. We observed elevated expression of floral repressor *LF* for in the domesticated *DTF2* allele, and a similar elevation in expression of the floral promoter *FTa1* for *DTF5*, consistent with the known roles of these genes and direction of the allelic effects of the two loci. Both loci potentially represent gain-of-function changes in the domesticated lineage. Analysis of the precise molecular basis for these QTL may require continued efforts to refine their map position and sequencing of the entire region surrounding FTa1/FTa2.

Chapter 8: General discussion

8.1 Summary of main findings

Wild and domesticated crops exhibit very distinct phenotypic differences that result from contrasting selective pressures and genetic isolation. This evolutionary transformation has often been considered to comprise two distinct phases; *domestication* and *diversification*. The *domestication* phase refers to the rapid modification of key traits incurred during an initial domestication episode (Ladizinsky, 1987, Ladizinsky, 1993, Abbo et al., 2011), followed by incremental adaptive modifications to non-essential traits, within the *diversification* phase. While there has been debate about this distinction (Abbo et al., 2014), it is clear that defining the genetic and molecular basis for these phenotypic changes will be key to a better understanding of the overall trajectory of domestication process.

The domesticated pea can be distinguished from wild forms by loss of seed dormancy and pod dehiscence, earlier flowering under short photoperiods, reduced branching and more robust growth habit, yet the genetic basis for these traits is still poorly understood. The genetics of pea domestication was last explored by Weeden (2007), however this study did not present a detailed QTL analysis, and appeared restricted by a low resolution genetic map, limiting the ability to clearly define the location of QTLs. Since that time there has been a rapid development of low-cost, high throughput genotyping technologies, which has increased efficiency and resolution of genetic mapping. As a consequence, medium to high resolution genetic maps have become available (Duarte et al., 2014, Sindhu et al., 2014, Sudheesh et al., 2015, Ma et al., 2017, Barilli et al., 2018) with the 13.2K consensus map constructed by Tayeh et al. (2015) being dramatic increase from the ~100 markers used by Weeden (2007).

In light of these developments, this study has re-examined in greater detail the underlying genetic control of pea domestication. This involved measuring and analysing a range of morphological and physiological traits typically associated with the domestication syndrome, in addition to a number of other traits not specifically related to domestication. A high-confidence and high-density linkage map was developed comprising of around 4600 markers, with QTL mapping conducted on a population of 136 individuals, a significantly larger population of those used in Weeden (2007). Our population was developed from a wide cross between modern cultivar *P. sativum var sativum* line NGB5839 and wild *P. sativum var humile* line JI1794*,* a representative of the "northern humile" lineage which is considered a major contributor to the modern day domesticated pea (Hoey et al., 1996, Zohary et al., 2012). An in-depth investigation and fine mapping was conducted on specific QTL regions related to two key domestication traits, seed dormancy and flowering time. Compared to previous work this study has significantly improved the mapping resolution of QTLs for a diverse range of domestication

related traits and appears to be the first study to systematically analyse these loci in advanced generation segregating populations.

Two major limitations in the scope of this study have been recognised. Firstly, because this study has been based on a single biparental cross, it is not clear that either parent is particularly representative of its type. Thus, we cannot yet be certain whether the QTLs discovered are widely represented in cultivated germplasm, although this is a reasonable assumption for the "major" domestication trait of seed dormancy. The second major limitation of this study is that the domesticated parent is a modern cultivar that carries not only the early, domestication-related changes but also many other changes accumulated during subsequent diversification and improvement. It is therefore not possible to determine when specific variations may have arisen, for example between wild and landrace and landrace to modern cultivar.

8.2 Overarching changes to genetic architecture

One of the findings to have emerged from this study was domestication-related traits were regulated by relatively few but major QTLs. This has been commonly reported not only in pea (Weeden, 2007) and other legumes (Koinange et al., 1996, Isemura et al., 2007, Liu et al., 2007, Isemura et al., 2010, Isemura et al., 2012, Kongjaimun et al., 2012) but is common among many crops (Paterson et al., 1995, Ross-Ibarra, 2005, Burger et al., 2008), which is likely driven by strong selective pressures from contrasting functional and environmental requirements. Interestingly, sunflower was an exception and instead exhibited a greater proportion of small to intermediate QTLs, but the reason for this is unknown (Burke et al., 2002, Wills and Burke, 2007).

The distribution of QTLs regulating growth habit and flowering time showed strong clustering to four genomic regions on LGs II, III and VI. QTL clustering has been reported in other crop models (Ross-Ibarra, 2005, Gepts, 2010) including in legume species Azuki bean (*Vigna angularis*) (Kaga et al., 2012) Rice bean (*Vigna umbellata*) (Isemura et al., 2010), Mungbean (Isemura et al., 2012) Yardlong bean and Cowpea (*Vigna unguiculata*) (Kongjaimun et al., 2012, Lo et al., 2018). However, this is the first-time clustering has been reported in pea, which contrasts with previous findings by Weeden (2007). Whether clustering is a consequence of pleiotropy or tightly linked genes is often debated and may differ in different situations. Pleiotropy has been shown in some cases, with certain traits having QTLs co-localising in multiple regions on the same genome; for example in mungbean, hard-seededness and seed size QTLs were mapped in close proximity on LGI, LGII and LGIII (Humphry et al., 2005, Isemura et al., 2012). In further support of pleiotropy, Lush and Evans (1980) reporting a correlation of reduced seed size with hardseededness in numerous other legume species, with QTLs also co-localised in other species such as tomato (Downie et al., 2003), arabidopsis (Debeaujon et al., 2000).

Interestingly co-localisation of seed size and hardseededness has also been shown in this thesis for pea on LGII. Here we suggest the co-localization of smaller seeds with increased seed dormancy reflects the competing pathways for carbon resources between the testa for fatty acid accumulation and the seed embryo for storage compounds (Baud et al., 2008, Chen et al., 2015, Li et al., 2018).

Many of the QTL clusters identified in this study were interpreted to likely reflect pleiotropic effects of flowering time genes, which are known to have a diverse range of functions (Pin and Nilsson, 2012, Tsuji, 2017). This was strongly supported by a comparison of near-isogenic lines for the *Hr* locus, and F_4 advanced generation segregating populations for *DTF6*/*E*, which showed plants carrying the earlier flowering domesticated allele also showed a range of growth habit and plant architecture traits typically associated with the domestication syndrome (Chapter 4). This is consistent with rapid evolution of domesticated crops (Hillman and Davies, 1990). In contrast, this study through advanced segregating populations, distinguished distinct QTLs within recognised QTL clusters, including seed coat roughness (*TT6a*/*GTY*) and a second testa thickness QTL (*TT6b*) within the Clst6 region on LGVI (Chapter 6). This indicates that not all QTLs can be explained by pleiotropy and in certain cases a more protracted evolution has occurred.

8.3 Seed dormancy related traits

In pea, loss of seed dormancy and seed dispersal mechanisms are widely acknowledged as critical domestication traits (Ladizinsky, 1987, Abbo et al., 2014, Hradilová et al., 2017). This has been demonstrated in cultivation experiments using wild pea, which showed poor crop establishment due to unpredictable germination rates and dramatic yield loss due to dehiscent pods (Abbo et al., 2011). While there have been recent advances into understanding the genetic control of indehiscent pods (Hradilová et al., 2017) dormancy loss remains poorly understood (Smykal et al., 2014) and consequently, this has been a major focus in this thesis.

The divergence of domesticated pea has typically been characterised by the loss of seed coat roughness, which is ubiquitous among cultivated lines (Zohary and Hopf, 1973, Zohary et al., 2012, Zaytseva et al., 2017), including the probable independent domesticate pea *P. sativum var abyssinicum* (Weeden, 2007). This trait has been used in archaeological studies as an indicator of domestication and implies some critical domestication function. A major locus contributing to the difference, *Gritty*, was first proposed by Marx (1969), and is known to be in a central region of LGVI (Weeden, 2007, Bordat et al., 2011). From our analysis, initial QTL mapping confirmed a major seed coat roughness QTL to the *GTY* locus, which we also found mapped near a testa thickness QTL (*TT6*). This providing a plausible but indirect link to seed dormancy given previous observations between testa thickness with hardseededness (Plitmann and Kislev, 1989, Miao et al., 2001, Weeden, 2007,

Weeden, 2018). However, despite its apparent significance, the map location of *GTY* remains poorly defined and its molecular nature unknown (Weeden and Wolko, 1990, Weeden, 2007). Using advanced generation segregating populations *GTY* was shown to co-segregate with testa thickness between anchor markers *CABB* and *FULα*. When extrapolated into the Tayeh et al. (2015) high density linkage map, this indicated a 6.4cM interval (51.1cM to 57.5cM), thereby providing the most refined mapping position of this locus. Surprisingly, permeability experiments identified no significant correlation between the *GTY* trait and permeability, thus indicating the primary reason for the selection of smooth-seeded genotypes was unrelated to decreased dormancy despite common perception. While it is conceivable that it may instead have been selected for improved palatability, potentially in part due to an association with reduced testa thickness, this does not seem sufficient to explain its fixation in domesticated material, particularly given its relatively small contribution to overall variance for testa thickness. Alternatively, it may have some other critical function not yet apparent.

Another surprising result from this study was the dissociation of testa thickness and permeability indicated by the identification of QTLs influencing one but not the other. Weeden (2018) proposed that the thinning of the seed coat during domestication was central to loss of dormancy, but results from this study suggested that only one of the six identified testa thickness QTLs had a significant effect on permeability. This implies a gradual thinning of the seed coat during domestication that may not necessarily linked to dormancy, therefore contradicting the suggestion by Murphy and Fuller (2017) that quantitative reduction in thickness was evidence for a protracted domestication in legumes.

In this study two permeability QTLs were identified, *Perm2* and *Perm7*. *Perm2* was considered as the critical domestication modification and therefore a focus of this study, due to its more significant effect (PEV = 32.7% compared to 8.6% for *Perm7*) and a similar QTL mapped within independent populations (Weeden, 2007). In comparison, there is no clear evidence of *Perm7* being reported in other populations and thus was subsequently recognised as a crop improvement modification. Fine mapping tentatively positioned *Perm2* between markers *NEK4* and *NAD* (4.4cM interval based on the Tayeh et al. (2015) linkage map), which included the well-known Mendel *A* gene, a *bHLH* transcription factor that is a strong positive regulator of flavonoid biosynthesis pathway (Hellens et al., 2010). The loss of seed pigmentation has been previously correlated with hardseededness in pea and a number of other legumes (Werker et al., 1979, Legesse and Powell, 1996, Souza and Marcos-Filho, 2001, Ross et al., 2010, Freixas Coutin et al., 2017). This is consistent with arabidopsis Mendel *A* orthologous gene *TT8* as well as other flavonoid biosynthesis pathway genes having pleiotropic effects on seed

dormancy and testa thickness (Debeaujon et al., 2000). If loss of pigmentation had been the key change permitting free germination and leading to domestication it would be expected that this trait would have undergone a severe selective sweep, but pigmented lines are widespread within the domesticated germplasm. One possible explanation for this may be that the causal gene for *Perm2* is merely closely linked to *A* and is not functionally related to pigmentation loss. A scan for other potential candidates within the *Perm2* region in Medicago identified four genes with a plausible connection to seed dormancy, through potential effects on the biosynthesis of hydroxylated fatty acids, a component strongly correlated with seed coat cracking and permeability (Shao et al., 2007, Chai et al., 2016, Cechová et al., 2017). However, this model again does not explain the absence of a genetic sweep in this region, with independent QTLs identified in separate populations (Weeden, 2007). Alternatively, dormancy loss may have evolved more than once, resulting from independent mechanisms. To examine this concept, genetic diversity data was extrapolated from Holdsworth et al. (2017) and combined with pigmentation mutant data from Hellens et al. (2010). Results showed no distinction in genetic diversity between lines exhibiting pigmented and non-pigmented phenotypes, therefore implying that if two mechanisms had co-evolved this did not result in independent evolutionary trajectories. Because there are currently no available induced Mendel *A* mutants on a wild/dormant background in pea, it is still disputable the role pigmentation loss has on dormancy.

Previous histological and chemical examination studies between dormant and non-dormant lines have highlighted thickening of palisade and cuticle layer, presence or absence of cracks and compositions of carbohydrates, hydroxylated fatty acids, or phenol compounds as potential contributors to testa permeability properties (Hradilová et al., 2017). While analysing these components has been outside the scope of this study subsequent investigations of the seed coat ultrastructure and chemical composition could reveal the nature of these two permeability QTLs. This could be used to explore our hypothesis that pigmentation loss caused increased permeability due to the reduction in hydroxylated fatty acid content. This model was based on the previous findings showing flavonoid biosynthesis genes *WD40* and *MYB* (*TTG1* and *TT2*) were involved in allocating of resources towards the seed coat for lipid biosynthesis (Baud et al., 2008, Chen et al., 2015, Li et al., 2018).

This study did not explore endogenous seed dormancy, as this has not been widely reported in *Fabaceae* family which is strongly modulated by the testa (Baskin and Baskin, 1998, Baskin, 2003). Interestingly, through the course of this thesis, varying germination rates of up to 30 days had been observed in scarified seeds in JI1794 x NGB5839 populations (although not reported) indicating possible, albeit minor role of endogenous dormancy in pea. While there have been a few species in this family to show endogenous physiological dormancy (Van Staden et al., 1987), there have been no

known reports in pea. It would therefore be interesting to investigate whether this observed variation has genetic component, and if so whether this was a domestication selective target.

8.4 Flowering time

Control of flowering has been widely recognised as an important adaptive mechanism for both wild and cultivated crops. While its genetic basis has been extensively studied in arabidopsis and many cereal crops it remains less understood in legumes, particularly from a domestication perspective. After its domestication in the Fertile Crescent, the cultivated pea underwent a rapid expansion along its latitudinal range, but was restricted longitudinally due to its obligate long-day photoperiod and cold tolerance requirements (Purugganan and Fuller, 2009, Jing et al., 2010). In a previous study using the same cross, two major QTLs for flowering time under short-days were identified. This included one on LGVI (QTL6) and a major QTL on LGIII encoded as the circadian clock-related *ELF3* gene, with the *elf3* mutation linked to the winter to spring cropping shift, enabling its eventual expansion into northern Europe (Weller et al., 2012). In this study, we revaluated the genetic control of flowering in long-day conditions which more closely reflects its normal growing conditions.

From the QTL analysis, five QTLs were identified; *DTF2, DTF3a, DTF3b, DTF5* and *DTF6*, with *DTF3a* (*ELF3a*) and *DTF6* (*QTL6*) coinciding with the two QTLs identified in short-day conditions (Weller et al., 2012). Based on similar phenotypic morphologies and corresponding map position to known induced mutant, *DTF3b* was considered highly likely to reflect a pleiotropic effect of the gibberellin biosynthesis gene *PsLe*, while *DTF6* mapped to a syntenic region in lupin, chickpea, Medicago and lentil which encompassed a novel *FT* homolog gene (Książkiewicz et al., 2016, Ortega, 2018), previously unreported in pea (Hecht et al., 2011, Laurie et al., 2011) but currently under investigation in a separate study. Subsequently this study concentrated on investigation of the of the novel QTLs, *DTF2* on LGII and *DTF5* on LGV.

From the collective evidence gathered from advanced generation segregating populations, synteny and expression analysis, *DTF5* was strongly linked to the *FTa1-FTa2-FTc* cluster and *DTF2* with *LATE FLOWERING* (*LF*) gene. Natural variation at the *LF* locus has been associated with differences in flowering time within the cultivated pea germplasm, (Murfet, 1975, Weller et al., 1997, Foucher et al., 2003, Weeden, 2007) and it is possible that both the NGB5839 and JI1794 alleles are present in the domesticated germplasm. However, this is the first time that natural variation for flowering time in the region of the LGV *FT* cluster has been identified in pea.

For *DTF2*, plants carrying the domesticated allele flowered later than those carrying the wild allele which was associated with an increase in *LF* expression. This is consistent with the known role of *LF* as

a repressor of flowering (Murfet, 1975, Foucher et al., 2003) and indicates that NGB5839 carries a gain-of-function allele, possibly interfering with a region in or around the *LF* gene that normally contributes to its repression. This is consistent with previous work that has suggested that regions outside the *LF* coding region are likely to play an important role in its regulation (Foucher et al., 2003). More detailed investigation and molecular characterisation of this was outside the scope of this study, but detailed sequence analysis of the *LF* gene and flanking sequences will likely be required.

In the case of *DTF5*, early flowering associated with the NGB5839 allele was also accompanied by increased expression of a candidate floral promoter gene *FTa1*. The important effects of this gene on flowering have already been demonstrated through study of induced loss-of-function mutants, and the effects of overexpression (Hecht et al., 2011, Laurie et al., 2011). Resolving the nature of this mutation was outside the scope of this study, but given apparent gain-of-function nature of the NGB5839 allele, the causal mutation(s) and mechanism might potentially be complex. Interestingly, the concurrent upregulation of the *FTa1-FTa2* intergenic region (*RMK*) in pea was consistent with findings implicating this region in flowering time control in other crop legumes, where QTL for flowering time have been identified in the corresponding chromosomal location. Dominant alleles from domesticated chickpea and low-latitude adapted lentil confer early flowering and show elevated expression of *FTa1* and other genes in the cluster (Rajandran, 2016, Ortega, 2018). Of particular interest are observations of upregulation in the *FTa1-FTa2* intergenic region in early lines, which has been associated with the vernalisation response in Medicago (Mauren et al., 2013, Yeoh et al., 2013, Ortega, 2018), and significant sequence differences affecting this region are present in chickpea and lentil (Rajandran, 2016, Ortega, 2018). Thus, the effect of *DTF5* could result from a similar mechanism. This warrants further investigation by conducting expression analysis on wild and domesticated lines under vernalisation and non-vernalisation conditions, and of course, by sequencing.

The results from this study thus validated flowering time QTLs on LGII and LGV, which were not identified in the former study (Weller et al., 2012). These QTLs were strongly linked to altered expression of *LF* and *FTa1*, however as previously mention further work is required to characterise the nature of these mutations. This would enable an investigation into the distribution and exploration of its evolutionary history by a sequence diversity study similar to that reported for the *DTF3*/*HR*/*ELF3a* gene by Weller et al. (2012). In addition, investigation of possible convergent evolution of these QTLs in other crop species could warrant further study. As previously mentioned the *FT* cluster has shown to be an important regulator of flowering in cultivated lentil (Rajandran, 2016) and chickpea (Ortega, 2018) while *ELF3a* mutants have been reported in lentil (Weller et al., 2012), chickpea (Ridge et al., 2017) and barley (Faure et al., 2012). It would therefore be interesting to discover if there is an

underlying genetic characteristic which makes these more predisposed for domestication than other flowering time genes.

8.5 Pod indehiscence

Although pod indehiscence was not reported in this study its fundamental role in cultivation has been widely reported. In pea, differences in dehiscence are mainly attributable to a single locus, *Dpo1*, which has clearly undergone a severe genetic sweep within the domesticated germplasm (Weeden et al., 2002, Weeden, 2007, Hradilová et al., 2017). Modifier genes for pod dehiscence have been reported but, unlike *Dpo1* these were not universally present across the germplasm (Weeden et al., 2002, Weeden, 2007), therefore implying later evolution. These additional QTLs predominately had a more minor effect than *dpo1* (Weeden et al., 2002) but in conjunction reinforced indehiscent phenotype in an additive manner. To put differently, the *dpo1* mutation may have provided the morphological change allowing domestication, whereas subsequent QTLs improved performance. Similar findings were reported in other crop models, such as soybean, where loss of pod dehiscence was universally attributed to the *shat1-5* transcription factor mutation (Dong et al., 2014, Sedivy et al., 2017), while a second QTL *pdh1* was not universally adopted (Funatsuki et al., 2014). This second mutation was found in a large proportion of Chinese landraces but absent in the Japanese, Korean, and Southeast Asian region landrace lines (Kaga et al., 2012). It was therefore proposed that *shat1-5* was sufficient in preventing shattering in more humid conditions, but a second mutation was required for the more arid environments (Funatsuki et al., 2014).

8.6 Genetic perspective to the domestication/diversification model

According to its definition, the *domestication* phase occurs between the transition of wild to landrace and involves the accumulation of critical modifications to the wild line to enable profitable cultivation practises, which from a morphological perspective in pea has been described as indehiscent pods and loss of seed dormancy (Abbo et al., 2011). This can be distinguished from the *diversification* phase which includes subsequent adaptive modifications to non-essential traits incurred between landrace and modern cultivar.

From a genetic perspective a critical domestication modification would likely involve a single major effecting gene to enable the plant to overcome restrictive barriers, and being preferentially selected would also show a severe genetic sweep across the domesticated germplasm. As illustrated in Figure 8.1, genetic evaluations of the critical domestication traits showed oligogenic control, indicating a near qualitative change (Abbo et al., 2011, Abbo et al., 2014), although additional modifying QTLs can also be present.

As discussed multiple QTLs have previously been identified for indehiscence pods, however in accordance to this genetic model only *Dpo1* can be considered as a critical domestication modification, whereas the others are considered diversification. This study looked at the genetic control of seed dormancy and through permeability experiments identified two QTLs, *Perm2* and *Perm7*. For reasons discussed in Section 8.3, *Perm2* was originally considered critical for the initial domestication step and *Perm7* arising later with additive effect. From fine mapping of *Perm2,* Mendel *A* was considered a strong candidate, however as discussed in Section 5.4.2 this mutation was not ubiquitous across the domesticated germplasm, therefore in accordance to the model this does not constitute as a critical domestication modification.

From a genetic viewpoint this model would also implicate loss of seed coat roughness as a domestication trait, as this shows both qualitative control and severe selective sweep across the domesticated germplasm. Interestingly our study found seed coat roughness had no significant correlation to permeability, and the functional significance of its selection is currently unknown. Whether this has some other critical function and/or effect on an additional and currently unrecognised domestication trait requires further study.

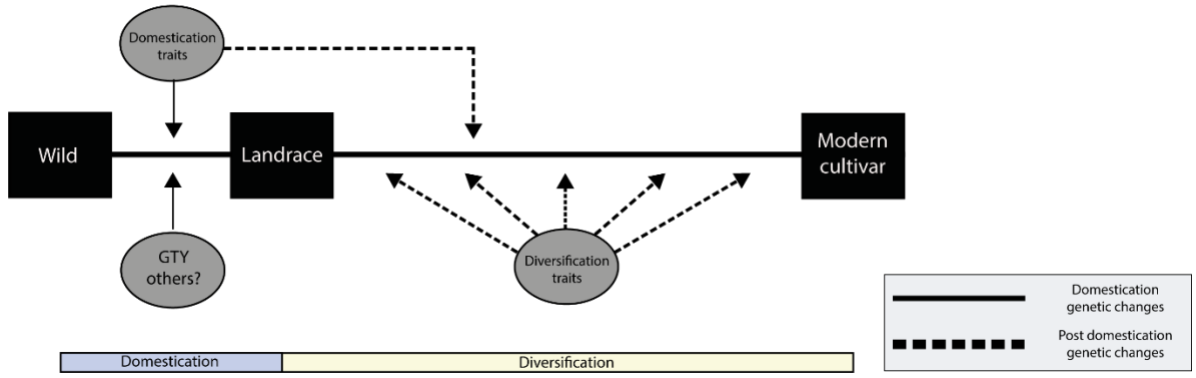


Figure 8. 1: Schematic representation the evolutionary modifications incurred during the domestication (blue box) and diversification (yellow box) phases. Domestication episode denoted by accumulation of critical domestication traits (loss of seed dormancy and indehiscent pods) which is regulated by a single mutation which occurred between wild and landrace. This includes additional modifications of unknown function, such as loss of seed coat roughness which are fixed across domesticated germplasm. Diversification phase includes accumulation of adaptive changes, including diversification traits and modifier effects domestication traits. This occurs between landrace and modern cultivar and can show more complex genetic regulation.

8.7 Future directions

Considering the complexity of the domestication process, the purpose of this project was not to provide a complete account of its genetic and molecular basis in pea, but instead offer new information about the genetic control of specific traits and contribute to the overall framework of knowledge on which to base future study.

As previously discussed, in focusing on only one cross and a single domesticated cultivar, this study accounts for only one of many potential evolutionary pathways. Future work should probably develop further the approach used by Weeden (2007) and expand the focus to several different comparative populations (wild x landrace, wild x modern cultivar and landrace x modern cultivar) to formulate a more comprehensive understanding. Using this approach, analysing the genetic control of dormancy between wild and more primitive, pigmented domesticated/landraces lines could help resolve questions arisen from this study, as to whether the *Perm2* locus is distinct from Mendel *A* gene. In future, association and diversity studies will also be informative, once they become feasible. Recently, the first such study in pea has shown selection of two QTLs between winter and spring types which conferred to known cold tolerance and flowering time regions (Siol et al., 2017).

Our ability to study crop domestication has improved significantly in recent years driven by new and emerging genetic tools and resources. For pea, however, a fully sequenced genome is currently unavailable but is imminent. As shown in this study, the access to a high-density linkage map as well as reference genomes of closely related species, including the recently released lentil genome, has dramatically improved ability to identify QTLs and explore candidate genes, relative to the last evaluation by Weeden (2007). However, this reliance on synteny in closely related species, can be limiting in poor syntenic regions, as shown for the mapping of testa thickness and seed coat roughness QTLs on LGVI. The release of the much-anticipated pea genome, will greatly improve ability to identify genes and functional elements as well as provide genomic tools for gene mapping, isolation and molecular breeding.

8.8 Concluding remarks

By comparing the genomic architecture of a representative wild and modern cultivar (JI1794 x NGB5839) the broad molecular implications of crop domestication have been discussed. Using QTL analysis domestication related genomic regions have been highlighted and two key domestication-related traits were genetically dissected using advanced generation segregating populations. By analysing the genetic architecture, mapping domestication related traits and fine mapping causative regions, this research has further contributed to our understanding to crop domestication. Most importantly this research has provided a platform for future studies which is transferable to in other legume crop models, while also providing a useful resource for plant breeders in future crop improvement efforts.

Bibliography

ABBO, S. & GOPHER, A. 2017. Near Eastern Plant Domestication: A History of Thought. *Trends in Plant Science,* 22**,** 491-511.

ABBO, S., PINHASI VAN-OSS, R., GOPHER, A., SARANGA, Y., OFNER, I. & PELEG, Z. 2014. Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. *Trends in Plant Science,* 19**,** 351-360.

ABBO, S., RACHAMIM, E., ZEHAVI, Y., ZEZAK, I., LEV-YADUN, S. & GOPHER, A. 2011. Experimental growing of wild pea in Israel and its bearing on Near Eastern plant domestication. *Annals of Botany,* 107**,** 1399-1404.

ABE, M., KOBAYASHI, Y., YAMAMOTO, S., DAIMON, Y., YAMAGUCHI, A., IKEDA, Y., ICHINOKI, H., NOTAGUCHI, M., GOTO, K. & ARAKI, T. 2005. FD, a bZIP Protein Mediating Signals from the Floral Pathway Integrator FT at the Shoot Apex. *Science,* 309**,** 1052-1056.

AGATI, G. & TATTINI, M. 2010. Multiple functional roles of flavonoids in photoprotection. *New Phytologist,* 186**,** 786-793.

ALBERT, S., DELSENY, M. & DEVIC, M. 1997. BANYULS, a novel negative regulator of flavonoid biosynthesis in the Arabidopsis seed coat. *The Plant Journal,* 11**,** 289-299.

ALI-RACHEDI, S., BOUINOT, D., WAGNER, M.-H., BONNET, M., SOTTA, B., GRAPPIN, P. & JULLIEN, M. 2004. Changes in endogenous abscisic acid levels during dormancy release and maintenance of mature seeds: studies with the Cape Verde Islands ecotype, the dormant model of Arabidopsis thaliana. *Planta,* 219**,** 479-488.

ALONSO-BLANCO, C., AARTS, M. G. M., BENTSINK, L., KEURENTJES, J. J. B., REYMOND, M., VREUGDENHIL, D. & KOORNNEEF, M. 2009. What Has Natural Variation Taught Us about Plant Development, Physiology, and Adaptation? *The Plant Cell,* 21**,** 1877-1896.

APPELHAGEN, I., NORDHOLT, N., SEIDEL, T., SPELT, K., KOES, R., QUATTROCHIO, F., SAGASSER, M. & WEISSHAAR, B. 2015. TRANSPARENT TESTA 13 is a tonoplast P3A-ATPase required for vacuolar deposition of proanthocyanidins in Arabidopsis thaliana seeds. *The Plant Journal,* 82**,** 840-849.

APPELHAGEN, I., THIEDIG, K., NORDHOLT, N., SCHMIDT, N., HUEP, G., SAGASSER, M. & WEISSHAAR, B. 2014. Update on transparent testa mutants from Arabidopsis thaliana: characterisation of new alleles from an isogenic collection. *Planta,* 240**,** 955-970.

ARGEL, P. & PATON, C. 1999. Overcoming Legume 14 hardseededness. *Forage seed production: tropical and subtropical species,* 2**,** 247-256.

AZANI, N., BABINEAU, M., BAILEY, C. D., BANKS, H., BARBOSA, A. R., PINTO, R. B., BOATWRIGHT, J. S., BORGES, L. M., BROWN, G. K., BRUNEAU, A., CANDIDO, E., CARDOSO, D., CHUNG, K. F., CLARK, R. P., CONCEICAO, A. D., CRISP, M., CUBAS, P., DELGADO-SALINAS, A., DEXTER, K. G., DOYLE, J. J., DUMINIL, J., EGAN, A. N., DE LA ESTRELLA, M., FALCAO, M. J., FILATOV, D. A., FORTUNA-PEREZ, A. P., FORTUNATO, R. H., GAGNON, E., GASSON, P., RANDO, J. G., TOZZI, A., GUNN, B., HARRIS, D., HASTON, E., HAWKINS, J. A., HERENDEEN, P. S., HUGHES, C. E., IGANCI, J. R. V., JAVADI, F., KANU, S. A., KAZEMPOUR-OSALOO, S., KITE, G. C., KLITGAARD, B. B., KOCHANOVSKI, F. J., KOENEN, E. J. M., KOVAR, L., LAVIN, M., LE ROUX, M., LEWIS, G. P., DE LIMA, H. C., LOPEZ-ROBERTS, M. C., MACKINDER, B., MAIA, V. H., MALECOT, V., MANSANO, V. F., MARAZZI, B., MATTAPHA, S., MILLER, J. T., MITSUYUKI, C., MOURA, T., MURPHY, D. J., NAGESWARA-RAO, M., NEVADO, B., NEVES, D., OJEDA, D. I., PENNINGTON, R. T., PRADO, D. E., PRENNER, G., DE QUEIROZ, L. P., RAMOS, G., FILARDI, F. L. R., RIBEIRO, P. G., RICO-ARCE, M. D., SANDERSON, M. J., SANTOS-SILVA, J., SAO-MATEUS, W. M. B., SILVA, M. J. S., SIMON, M. F., SINOU, C., SNAK, C., DE SOUZA, E. R., SPRENT, J., STEELE, K. P., STEIER, J. E., STEEVES, R., STIRTON, C. H., TAGANE, S., TORKE, B. M., TOYAMA, H., DA CRUZ, D. T., VATANPARAST, M., WIERINGA, J. J., WINK, M., WOJCIECHOWSKI, M. F., YAHARA, T., YI, T. S.

& ZIMMERMAN, E. 2017. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. *Taxon,* 66**,** 44-77.
BAJAJ, D., DAS, S., UPADHYAYA, H. D., RANJAN, R., BADONI, S., KUMAR, V., TRIPATHI, S., GOWDA, C. L. L., SHARMA, S., SINGH, S., TYAGI, A. K. & PARIDA, S. K. 2015. A Genome-wide Combinatorial Strategy Dissects Complex Genetic Architecture of Seed Coat Color in Chickpea. *Frontiers in Plant Science,* 6**,** 979.
BARBER, H. 1959. Physiological genetics of Pisum II. *Heredity,* 13**,** 33.
BARILLI, E., COBOS, M. J., CARRILLO, E., KILIAN, A., CARLING, J. & RUBIALES, D. 2018. A High-Density Integrated DArTseq SNP-Based Genetic Map of Pisum fulvum and Identification of QTLs Controlling Rust Resistance. *Frontiers in Plant Science,* 9**,** 13.
BASKIN, C. C. 2003. Breaking physical dormancy in seeds–focussing on the lens. *New Phytologist,* 158**,** 229-232.
BASKIN, C. C. & BASKIN, J. M. 1998. *Seeds: ecology, biogeography, and, evolution of dormancy and germination*, Elsevier.
BASKIN, J. M. & BASKIN, C. C. 2000. Evolutionary considerations of claims for physical dormancy-break by microbial action and abrasion by soil particles. *Seed Science Research,* 10**,** 409-413.
BASKIN, J. M. & BASKIN, C. C. 2007. A classification system for seed dormancy. *Seed Science Research,* 14**,** 1-16.
BASKIN, J. M., BASKIN, C. C. & LI, X. 2000. Taxonomy, anatomy and evolution of physical dormancy in seeds. *Plant Species Biology,* 15**,** 139-152.
BAUD, S., BOUTIN, J.-P., MIQUEL, M., LEPINIEC, L. & ROCHAT, C. 2002. An integrated overview of seed development in Arabidopsis thaliana ecotype WS. *Plant Physiology and Biochemistry,* 40**,** 151-160.
BAUD, S., DUBREUCQ, B., MIQUEL, M., ROCHAT, C. & LEPINIEC, L. 2008. Storage Reserve Accumulation in Arabidopsis: Metabolic and Developmental Control of Seed Filling. *The Arabidopsis Book / American Society of Plant Biologists,* 6**,** e0113.
BAUDRY, A., HEIM, M. A., DUBREUCQ, B., CABOCHE, M., WEISSHAAR, B. & LEPINIEC, L. 2004. TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in Arabidopsis thaliana. *The Plant Journal,* 39**,** 366-380.
BEISSON, F., LI, Y., BONAVENTURE, G., POLLARD, M. & OHLROGGE, J. B. 2007. The Acyltransferase GPAT5 Is Required for the Synthesis of Suberin in Seed Coat and Root of Arabidopsis. *Plant Cell,* 19**,** 351-68.
BEISSON, F., LI-BEISSON, Y. & POLLARD, M. 2012. Solving the puzzles of cutin and suberin polymer biosynthesis. *Current opinion in plant biology,* 15**,** 329-337.
BEN ZE'EV, N. & ZOHARY, D. 1973. Species relationships in the genus Pisum L. *Israel J. Bot,* 22**,** 73-91.
BENDIX, C., MARSHALL, CARINE M. & HARMON, FRANK G. 2015. Circadian Clock Genes Universally Control Key Agricultural Traits. *Molecular Plant,* 8**,** 1135-1152.
BENECH-ARNOLD, R. L. & RODRÍGUEZ, M. V. 2018. Pre-harvest Sprouting and Grain Dormancy in Sorghum bicolor: What Have We Learned? *Front Plant Sci,* 9.
BENNETT MD, L. I. 2012. *Angiosperm DNA C-values database* [Online]. Available: http://www.kew.org/cvalues/ [Accessed].
BERTIOLI, D. J., CANNON, S. B., FROENICKE, L., HUANG, G., FARMER, A. D., CANNON, E. K. S., LIU, X., GAO, D., CLEVENGER, J., DASH, S., REN, L., MORETZSOHN, M. C., SHIRASAWA, K., HUANG, W., VIDIGAL, B., ABERNATHY, B., CHU, Y., NIEDERHUTH, C. E., UMALE, P., ARAÚJO, A. C. G., KOZIK, A., DO KIM, K., BUROW, M. D., VARSHNEY, R. K., WANG, X., ZHANG, X., BARKLEY, N., GUIMARÃES, P. M., ISOBE, S., GUO, B., LIAO, B., STALKER, H. T., SCHMITZ, R. J., SCHEFFLER, B. E., LEAL-BERTIOLI, S. C. M., XUN, X., JACKSON, S. A., MICHELMORE, R. & OZIAS-AKINS, P. 2016. The genome sequences of Arachis duranensis and Arachis ipaensis, the diploid ancestors of cultivated peanut. *Nature Genetics,* 48**,** 438.
BEVERIDGE, C. A. 2000. Long-distance signalling and a mutational analysis of branching in pea. *Plant Growth Regulation,* 32**,** 193-203.

BEVERIDGE, C. A. & MURFET, I. C. 1996. The gigas mutant in pea is deficient in the floral stimulus. *Physiologia Plantarum,* 96**,** 637-645.

BEWLEY, J. D. 1997. Seed germination and dormancy. *Plant Cell,* 9**,** 1055-1066.

BHAKTA, M. S., GEZAN, S. A., MICHELANGELI, J. A. C., CARVALHO, M., ZHANG, L., JONES, J. W., BOOTE, K. J., CORRELL, M. J., BEAVER, J., OSORNO, J. M., COLBERT, R., RAO, I., BEEBE, S., GONZALEZ, A., RICAURTE, J. & VALLEJOS, C. E. 2017. A Predictive Model for Time-to-Flowering in the Common Bean Based on QTL and Environmental Variables. *G3-Genes Genomes Genetics,* 7**,** 3901-3912.

BLACKMAN, B. K. 2013. Interacting duplications, fluctuating selection, and convergence: the complex dynamics of flowering time evolution during sunflower domestication. *Journal of Experimental Botany,* 64**,** 421-431.

BLACKMAN, B. K., RASMUSSEN, D. A., STRASBURG, J. L., RADUSKI, A. R., BURKE, J. M., KNAPP, S. J., MICHAELS, S. D. & RIESEBERG, L. H. 2011. Contributions of Flowering Time Genes to Sunflower Domestication and Improvement. *Genetics,* 187**,** 271-87.

BLAIR, M. W., CORTÉS, A. J., FARMER, A. D., HUANG, W., AMBACHEW, D., PENMETSA, R. V., CARRASQUILLA-GARCIA, N., ASSEFA, T. & CANNON, S. B. 2018. Uneven recombination rate and linkage disequilibrium across a reference SNP map for common bean (Phaseolus vulgaris L.). *PloS one,* 13**,** e0189597.

BLIXT, S. 1972. Mutation genetics in Pisum. *Agri Hortique Genetica,* 30.

BOERJAN, W., RALPH, J. & BAUCHER, M. 2003. Lignin biosynthesis. *Annual review of plant biology,* 54**,** 519-546.

BOGDANOVA, V. S., KOSTERIN, O. E. & YADRIKHINSKIY, A. K. 2014. Wild peas vary in their cross-compatibility with cultivated pea (Pisum sativum subsp. sativum L.) depending on alleles of a nuclear–cytoplasmic incompatibility locus. *Theoretical and Applied Genetics,* 127**,** 1163-1172.

BÖHLENIUS, H., HUANG, T., CHARBONNEL-CAMPAA, L., BRUNNER, A. M., JANSSON, S., STRAUSS, S. H. & NILSSON, O. 2006. *CO/FT* Regulatory Module Controls Timing of Flowering and Seasonal Growth Cessation in Trees. *Science,* 312**,** 1040-1043.

BOMBLIES, K. & DOEBLEY, J. F. 2006. Pleiotropic effects of the duplicate maize FLORICAULA/LEAFY genes zfl1 and zfl2 on traits under selection during maize domestication. *Genetics,* 172**,** 519-531.

BORDAT, A., SAVOIS, V., NICOLAS, M., SALSE, J., CHAUVEAU, A., BOURGEOIS, M., POTIER, J., HOUTIN, H., ROND, C., MURAT, F., MARGET, P., AUBERT, G. & BURSTIN, J. 2011. Translational Genomics in Legumes Allowed Placing *In Silico* 5460 Unigenes on the Pea Functional Map and Identified Candidate Genes in *Pisum sativum L*. *G3: Genes|Genomes|Genetics,* 1**,** 93-103.

BOSS, P. K., BASTOW, R. M., MYLNE, J. S. & DEAN, C. 2004. Multiple pathways in the decision to flower: enabling, promoting, and resetting. *The Plant Cell,* 16**,** S18-S31.

BRUNEAU, A., DOYLE, J. J., HERENDEEN, P., HUGHES, C., KENICER, G., LEWIS, G., MACKINDER, B., PENNINGTON, R. T., SANDERSON, M. J., WOJCIECHOWSKI, M. F., BOATWRIGHT, S., BROWN, G., CARDOSO, D., CRISP, M., EGAN, A., FORTUNATO, R. H., HAWKINS, J., KAJITA, T., KLITGAARD, B., KOENEN, E., LAVIN, M., LUCKOW, M., MARAZZI, B., MCMAHON, M. M., MILLER, J. T., MURPHY, D. J., OHASHI, H., DE QUEIROZ, L. P., RICO, L., SARKINEN, T., SCHRIRE, B., SIMON, M. F., SOUZA, E. R., STEELE, K., TORKE, B. M., WIERINGA, J. J., VAN WYK, B. E. & LEGUME PHYLOGENY WORKING, G. 2013. Legume phylogeny and classification in the 21st century: Progress, prospects and lessons for other species-rich clades. *Taxon,* 62**,** 217-248.

BRUNETTI, C., DI FERDINANDO, M., FINI, A., POLLASTRI, S. & TATTINI, M. 2013. Flavonoids as antioxidants and developmental regulators: relative significance in plants and humans. *International journal of molecular sciences,* 14**,** 3540-3555.

BULMER, M. G. 1984. *Delayed germination of seeds: Cohen's model revisited*.

BURGER, J. C., CHAPMAN, M. A. & BURKE, J. M. 2008. Molecular insights into the evolution of crop plants. *American Journal of Botany,* 95**,** 113-122.

BURKE, J. M., TANG, S., KNAPP, S. J. & RIESEBERG, L. H. 2002. Genetic Analysis of Sunflower Domestication. *Genetics,* 161**,** 1257-1267.

BUTLER, J. B., VAILLANCOURT, R. E., POTTS, B. M., LEE, D. J., KING, G. J., BATEN, A., SHEPHERD, M. & FREEMAN, J. S. 2017. Comparative genomics of Eucalyptus and Corymbia reveals low rates of genome structural rearrangement. *Bmc Genomics,* 18**,** 13.

CAICEDO, A. L., STINCHCOMBE, J. R., OLSEN, K. M., SCHMITT, J. & PURUGGANAN, M. D. 2004. Epistatic interaction between *Arabidopsis FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 15670-15675.

CAVANAGH, C. R., CHAO, S., WANG, S., HUANG, B. E., STEPHEN, S., KIANI, S., FORREST, K., SAINTENAC, C., BROWN-GUEDIRA, G. L. & AKHUNOVA, A. 2013. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proceedings of the national academy of sciences,* 110**,** 8057-8062.

CECHOVÁ, M., VÁLKOVÁ, M., HRADILOVÁ, I., JANSKÁ, A., SOUKUP, A., SMÝKAL, P. & BEDNÁŘ, P. 2017. Towards Better Understanding of Pea Seed Dormancy Using Laser Desorption/Ionization Mass Spectrometry. *International journal of molecular sciences,* 18**,** 2196.

CHAI, M. F., ZHOU, C. N., MOLINA, I., FU, C. X., NAKASHIMA, J., LI, G. F., ZHANG, W. Z., PARK, J. J., TANG, Y. H., JIANG, Q. Z. & WANG, Z. Y. 2016. A class II KNOX gene, KNOX4, controls seed physical dormancy. *Proceedings of the National Academy of Sciences of the United States of America,* 113**,** 6997-7002.

CHEN, M., XUAN, L., WANG, Z., ZHOU, L., LI, Z., DU, X., ALI, E., ZHANG, G. & JIANG, L. 2014. TRANSPARENT TESTA8 Inhibits Seed Fatty Acid Accumulation by Targeting Several Seed Development Regulators in Arabidopsis. *Plant Physiology,* 165**,** 905-916.

CHEN, M., ZHANG, B., LI, C., KULAVEERASINGAM, H., CHEW, F. T. & YU, H. 2015. TRANSPARENT TESTA GLABRA 1 regulates the accumulation of seed storage reserves in Arabidopsis. *Plant physiology***,** pp. 00943.2015.

CHEN, X., LI, H., PANDEY, M. K., YANG, Q., WANG, X., GARG, V., LI, H., CHI, X., DODDAMANI, D., HONG, Y., UPADHYAYA, H., GUO, H., KHAN, A. W., ZHU, F., ZHANG, X., PAN, L., PIERCE, G. J., ZHOU, G., KRISHNAMOHAN, K. A. V. S., CHEN, M., ZHONG, N., AGARWAL, G., LI, S., CHITIKINENI, A., ZHANG, G.-Q., SHARMA, S., CHEN, N., LIU, H., JANILA, P., LI, S., WANG, M., WANG, T., SUN, J., LI, X., LI, C., WANG, M., YU, L., WEN, S., SINGH, S., YANG, Z., ZHAO, J., ZHANG, C., YU, Y., BI, J., ZHANG, X., LIU, Z.-J., PATERSON, A. H., WANG, S., LIANG, X., VARSHNEY, R. K. & YU, S. 2016. Draft genome of the peanut A-genome progenitor (*Arachis duranensis*) provides insights into geocarpy, oil biosynthesis, and allergens. *Proceedings of the National Academy of Sciences,* 113**,** 6785-6790.

CHILDS, D. Z., METCALF, C. J. E. & REES, M. 2010. Evolutionary bet-hedging in the real world: empirical evidence and challenges revealed by plants. *Proceedings of the Royal Society B: Biological Sciences*.

CHO, Y. B., JONES, S. I. & VODKIN, L. O. 2017. Mutations in Argonaute5 Illuminate Epistatic Interactions of the K1 and I Loci Leading to Saddle Seed Color Patterns in Glycine max. *The Plant Cell,* 29**,** 708-725.

CHOI, H.-K., MUN, J.-H., KIM, D.-J., ZHU, H., BAEK, J.-M., MUDGE, J., ROE, B., ELLIS, N., DOYLE, J., KISS, G. B., YOUNG, N. D. & COOK, D. R. 2004. Estimating genome conservation between crop and model legume species. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 15289-15294.

CHONO, M., HONDA, I., SHINODA, S., KUSHIRO, T., KAMIYA, Y., NAMBARA, E., KAWAKAMI, N., KANEKO, S. & WATANABE, Y. 2006. Field studies on the regulation of abscisic acid content

and germinability during grain development of barley: molecular and chemical analysis of pre-harvest sprouting. *Journal of Experimental Botany,* 57**,** 2421-2434.

CLARK, R. M., WAGLER, T. N., QUIJADA, P. & DOEBLEY, J. 2006. A distant upstream enhancer at the maize domestication gene tb1 has pleiotropic effects on plant and inflorescent architecture. *Nature genetics,* 38**,** 594.

CLOSE, T. J., BHAT, P. R., LONARDI, S., WU, Y., ROSTOKS, N., RAMSAY, L., DRUKA, A., STEIN, N., SVENSSON, J. T., WANAMAKER, S., BOZDAG, S., ROOSE, M. L., MOSCOU, M. J., CHAO, S., VARSHNEY, R. K., SZŰCS, P., SATO, K., HAYES, P. M., MATTHEWS, D. E., KLEINHOFS, A., MUEHLBAUER, G. J., DEYOUNG, J., MARSHALL, D. F., MADISHETTY, K., FENTON, R. D., CONDAMINE, P., GRANER, A. & WAUGH, R. 2009. Development and implementation of high-throughput SNP genotyping in barley. *BMC Genomics,* 10**,** 582.

COHEN, D. 1966. Optimizing reproduction in a randomly varying environment. *Journal of Theoretical Biology,* 12**,** 119-129.

COLLARD, B., MACE, E., MCPHAIL, M., WENZL, P., CAKIR, M., FOX, G., POULSEN, D. & JORDAN, D. 2009. How accurate are the marker orders in crop linkage maps generated from large marker datasets? *Crop and Pasture Science,* 60**,** 362-372.

COMADRAN, J., KILIAN, B., RUSSELL, J., RAMSAY, L., STEIN, N., GANAL, M., SHAW, P., BAYER, M., THOMAS, W., MARSHALL, D., HEDLEY, P., TONDELLI, A., PECCHIONI, N., FRANCIA, E., KORZUN, V., WALTHER, A. & WAUGH, R. 2012. Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. *Nature Genetics,* 44**,** 1388.

COOPER, J. W., WILSON, M. H., DERKS, M. F. L., SMIT, S., KUNERT, K. J., CULLIS, C. & FOYER, C. H. 2017. Enhancing faba bean (Vicia faba L.) genome resources. *J Exp Bot,* 68**,** 1941-53.

DANILEVSKAYA, O. N., MENG, X., HOU, Z., ANANIEV, E. V. & SIMMONS, C. R. 2008. A Genomic and Expression Compendium of the Expanded *PEBP* Gene Family from Maize. *Plant Physiology,* 146**,** 250-264.

DARVASI, A., WEINREB, A., MINKE, V., WELLER, J. I. & SOLLER, M. 1993. Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics,* 134**,** 943-951.

DE GIORGI, J., PISKUREWICZ, U., LOUBERY, S., UTZ-PUGIN, A., BAILLY, C., MENE-SAFFRANE, L. & LOPEZ-MOLINA, L. 2015. An Endosperm-Associated Cuticle Is Required for Arabidopsis Seed Viability, Dormancy and Early Control of Germination. *Plos Genetics,* 11.

DE VEGA, J. J., AYLING, S., HEGARTY, M., KUDRNA, D., GOICOECHEA, J. L., ERGON, Å., ROGNLI, O. A., JONES, C., SWAIN, M., GEURTS, R., LANG, C., MAYER, K. F. X., RÖSSNER, S., YATES, S., WEBB, K. J., DONNISON, I. S., OLDROYD, G. E. D., WING, R. A., CACCAMO, M., POWELL, W., ABBERTON, M. T. & SKØT, L. 2015. Red clover (Trifolium pratense L.) draft genome provides a platform for trait improvement. *Scientific Reports,* 5**,** 17394.

DEBEAUJON, I., LEON-KLOOSTERZIEL, K. M. & KOORNNEEF, M. 2000. Influence of the testa on seed dormancy, germination, and longevity in Arabidopsis. *Plant Physiology,* 122**,** 403-413.

DEBEAUJON, I., PEETERS, A. J., LÉON-KLOOSTERZIEL, K. M. & KOORNNEEF, M. 2001. The TRANSPARENT TESTA12 gene of Arabidopsis encodes a multidrug secondary transporter-like protein required for flavonoid sequestration in vacuoles of the seed coat endothelium. *The Plant Cell,* 13**,** 853-871.

DEMASON, D. A. & WEEDEN, N. F. 2006. Two Argonaute1 genes from pea. *Pisum Genetics,* 38**,** 3-9.

DÍAZ, A. M., CALDAS, G. V. & BLAIR, M. W. 2010. Concentrations of condensed tannins and anthocyanins in common bean seed coats. *Food Research International,* 43**,** 595-601.

DODD, A. N., SALATHIA, N., HALL, A., KÉVEI, E., TÓTH, R., NAGY, F., HIBBERD, J. M., MILLAR, A. J. & WEBB, A. A. R. 2005. Plant Circadian Clocks Increase Photosynthesis, Growth, Survival, and Competitive Advantage. *Science,* 309**,** 630-633.

DOEBLEY, J., STEC, A. & GUSTUS, C. 1995. teosinte branched1 and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics,* 141**,** 333-346.

DOEBLEY, J., STEC, A. & HUBBARD, L. 1997. The evolution of apical dominance in maize. *Nature,* 386**,** 485.
DOEBLEY, J. F., GAUT, B. S. & SMITH, B. D. 2006. The Molecular Genetics of Crop Domestication. *Cell,* 127**,** 1309-1321.
DONALD, C. M. 1968. The breeding of crop ideotypes. *Euphytica,* 17**,** 385-403.
DONG, Y., YANG, X., LIU, J., WANG, B.-H., LIU, B.-L. & WANG, Y.-Z. 2014. Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean. *Nature Communications,* 5**,** 3352.
DOSS, R. P., PROEBSTING, W. M., POTTER, S. W. & CLEMENT, S. L. 1995. Response ofNp mutant of pea (Pisum sativum L.) to pea weevil (Bruchus pisorum L.) oviposition and extracts. *Journal of chemical ecology,* 21**,** 97-106.
DOWNIE, A. B., ZHANG, D., DIRK, L. M., THACKER, R. R., PFEIFFER, J. A., DRAKE, J. L., LEVY, A. A., BUTTERFIELD, D. A., BUXTON, J. W. & SNYDER, J. C. 2003. Communication between the maternal testa and the embryo and/or endosperm affect testa attributes in tomato. *Plant physiology,* 133**,** 145-160.
DUARTE, J., RIVIERE, N., BARANGER, A., AUBERT, G., BURSTIN, J., CORNET, L., LAVAUD, C., LEJEUNE-HENAUT, I., MARTINANT, J. P., PICHON, J. P., PILET-NAYEL, M. L. & BOUTET, G. 2014. Transcriptome sequencing for high throughput SNP development and genetic mapping in Pea. *Bmc Genomics,* 15**,** 15.
DURAND, E., BOUCHET, S., BERTIN, P., RESSAYRE, A., JAMIN, P., CHARCOSSET, A., DILLMANN, C. & TENAILLON, M. I. 2012. Flowering time in maize: linkage and epistasis at a major effect locus. *Genetics***,** genetics. 111.136903.
DURANTI, M. & GIUS, C. 1997. Legume seeds: protein content and nutritional value. *Field Crops Research,* 53**,** 31-45.
EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research,* 32**,** 1792-1797.
ELLIS, T. & POYSER, S. 2002. An integrated and comparative view of pea genetic and cytogenetic maps. *New Phytologist,* 153**,** 17-25.
ELLIS, T. H. N., HELLENS, R. P., TURNER, L., LEE, C., DOMONEY, C. & WELHAM, T. 1993. On the pea linkage map. *Pisum Genetics,* 25**,** 5-12.
ELLIS, T. H. N., POYSER, S. J., KNOX, M. R., VERSHININ, A. V. & AMBROSE, M. J. 1998. Polymorphism of insertion sites of Ty1-copia class retrotransposons and its use for linkage and diversity analysis in pea. *Molecular and General Genetics,* 260**,** 9-19.
ESPINOZA, L. D. C. L., HUGUET, T. & JULIER, B. 2012. Multi-population QTL detection for aerial morphogenetic traits in the model legume Medicago truncatula. *Theoretical and Applied Genetics,* 124**,** 739-754.
FAOSTAT, F. 2018. *Food and Agricultural Organization of the United Nations* [Online]. Available: http://www.fao.org/faostat/en [Accessed 29/06/18].
FAURE, S., TURNER, A. S., GRUSZKA, D., CHRISTODOULOU, V., DAVIS, S. J., VON KORFF, M. & LAURIE, D. A. 2012. Mutation at the circadian clock gene *EARLY MATURITY 8* adapts domesticated barley (*Hordeum vulgare*) to short growing seasons. *Proceedings of the National Academy of Sciences,* 109**,** 8328-8333.
FERRARO, K., JIN, A. L., NGUYEN, T.-D., REINECKE, D. M., OZGA, J. A. & RO, D.-K. 2014. Characterization of proanthocyanidin metabolism in pea (Pisum sativum) seeds. *BMC Plant Biology,* 14**,** 238.
FERRER, J. L., AUSTIN, M. B., STEWART, C. & NOEL, J. P. 2008. Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. *Plant Physiology and Biochemistry,* 46**,** 356-370.
FINCH-SAVAGE, W. E. & FOOTITT, S. 2012. To germinate or not to germinate: a question of dormancy relief not germination stimulation. *Seed Science Research,* 22**,** 243-248.

FINKELSTEIN, R., REEVES, W., ARIIZUMI, T. & STEBER, C. 2008. Molecular aspects of seed dormancy. *Annual Review of Plant Biology.*

FOLEY, M. E. 2001. Seed dormancy: an update on terminology, physiological genetics, and quantitative trait loci regulating germinability. *Weed Science,* 49**,** 305-317.

FOUCHER, F., MORIN, J., COURTIADE, J., CADIOUX, S., ELLIS, N., BANFIELD, M. J. & RAMEAU, C. 2003. *DETERMINATE* and *LATE FLOWERING* Are Two *TERMINAL FLOWER1/CENTRORADIALIS* Homologs That Control Two Distinct Phases of Flowering Initiation and Development in Pea. *The Plant Cell,* 15**,** 2742-2754.

FRANKLIN, K. A. & QUAIL, P. H. 2010. Phytochrome functions in Arabidopsis development. *Journal of Experimental Botany,* 61**,** 11-24.

FREIXAS COUTIN, J. A., MUNHOLLAND, S., SILVA, A., SUBEDI, S., LUKENS, L., CROSBY, W. L., PAULS, K. P. & BOZZO, G. G. 2017. Proanthocyanidin accumulation and transcriptional responses in the seed coat of cranberry beans (Phaseolus vulgaris L.) with different susceptibility to postharvest darkening. *BMC Plant Biology,* 17**,** 89.

FUDGE, J. B., LEE, R. H., LAURIE, R. E., MYSORE, K. S., WEN, J., WELLER, J. L. & MACKNIGHT, R. C. 2018. Medicago truncatula SOC1 Genes Are Up-regulated by Environmental Cues That Promote Flowering. *Frontiers in Plant Science,* 9.

FULLER, D. Q. & ALLABY, R. 2018. Seed dispersal and crop domestication: shattering, germination and seasonality in evolution under cultivation. *Annual Plant Reviews online***,** 238-295.

FULLER, D. Q., DENHAM, T., ARROYO-KALIN, M., LUCAS, L., STEVENS, C. J., QIN, L., ALLABY, R. G. & PURUGGANAN, M. D. 2014. Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record. *Proceedings of the National Academy of Sciences of the United States of America,* 111**,** 6147-6152.

FULLER, D. Q., WILLCOX, G. & ALLABY, R. G. 2011. Cultivation and domestication had multiple origins: arguments against the core area hypothesis for the origins of agriculture in the Near East. *World Archaeology,* 43**,** 628-652.

FULLER, D. Q., WILLCOX, G. & ALLABY, R. G. 2012. Early agricultural pathways: moving outside the 'core area' hypothesis in Southwest Asia. *Journal of Experimental Botany,* 63**,** 617-633.

FUNATSUKI, H., SUZUKI, M., HIROSE, A., INABA, H., YAMADA, T., HAJIKA, M., KOMATSU, K., KATAYAMA, T., SAYAMA, T., ISHIMOTO, M. & FUJINO, K. 2014. Molecular basis of a shattering resistance boosting global dissemination of soybean. *Proceedings of the National Academy of Sciences,* 111**,** 17797-17802.

GAO, P., LI, X., CUI, D., WU, L., PARKIN, I. & GRUBER, M. Y. 2010. A new dominant Arabidopsis transparent testa mutant, sk21-D, and modulation of seed flavonoid biosynthesis by KAN4. *Plant Biotechnology Journal,* 8**,** 979-993.

GARCIA, D., FITZ GERALD, J. N. & BERGER, F. 2005. Maternal Control of Integument Cell Elongation and Zygotic Control of Endosperm Growth Are Coordinated to Determine Seed Size in Arabidopsis. *The Plant Cell,* 17**,** 52-60.

GAUR, P. M., SAMINENI, S., TRIPATHI, S., VARSHNEY, R. K. & GOWDA, C. L. L. 2015. Allelic relationships of flowering time genes in chickpea. *Euphytica,* 203**,** 295-308.

GELETA, M. & ORTIZ, R. 2016. Chapter Five - Molecular and Genomic Tools Provide Insights on Crop Domestication and Evolution. *In:* DONALD, L. S. (ed.) *Advances in Agronomy.* Academic Press.

GEPTS, P. 2010. Crop domestication as a long-term selection experiment. *Plant breeding reviews,* 24**,** 1-44.

GEPTS, P. 2014. The contribution of genetic and genomic approaches to plant domestication studies. *Current opinion in plant biology,* 18**,** 51-59.

GIRARD, A. L., BEAN, S. R., TILLEY, M., ADRIANOS, S. L. & AWIKA, J. M. 2018. Interaction mechanisms of condensed tannins (proanthocyanidins) with wheat gluten proteins. *Food Chemistry,* 245**,** 1154-1162.

GOFF, S. A., RICKE, D., LAN, T.-H., PRESTING, G., WANG, R., DUNN, M., GLAZEBROOK, J., SESSIONS, A., OELLER, P. & VARMA, H. 2002. A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). *Science,* 296**,** 92-100.

GONZALEZ, A., BROWN, M., HATLESTAD, G., AKHAVAN, N., SMITH, T., HEMBD, A., MOORE, J., MONTES, D., MOSLEY, T., RESENDEZ, J., NGUYEN, H., WILSON, L., CAMPBELL, A., SUDARSHAN, D. & LLOYD, A. 2016a. TTG2 controls the developmental regulation of seed coat tannins in Arabidopsis by regulating vacuolar transport steps in the proanthocyanidin pathway. *Developmental Biology,* 419**,** 54-63.

GONZALEZ, A. M., YUSTE-LISBONA, F. J., SABURIDO, S., BRETONES, S., DE RON, A. M., LOZANO, R. & SANTALLA, M. 2016b. Major Contribution of Flowering Time and Vegetative Growth to Plant Production in Common Bean As Deduced from a Comparative Genetic Mapping. *Frontiers in Plant Science,* 7**,** 18.

GOODSTEIN, D. M., SHU, S., HOWSON, R., NEUPANE, R., HAYES, R. D., FAZO, J., MITROS, T., DIRKS, W., HELLSTEN, U., PUTNAM, N. & ROKHSAR, D. S. 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research,* 40**,** D1178-D1186.

GOU, M. Y., HOU, G. C., YANG, H. J., ZHANG, X. B., CAI, Y. H., KAI, G. Y. & LIU, C. J. 2017. The MYB107 Transcription Factor Positively Regulates Suberin Biosynthesis. *Plant Physiology,* 173**,** 1045-1058.

GOULD, K. S., YOUNG, J. P. W. & CUTTER, E. G. 1992. L-system analysis of compound leaf development in Pisum sativum L. *Annals of Botany,* 70**,** 189-196.

GREILHUBER, J. & EBERT, I. 1994. Genome size variation in Pisum sativum. *Genome,* 37**,** 646-655.

GROOT, S. & KARSSEN, C. 1987. Gibberellins regulate seed germination in tomato by endosperm weakening: a study with gibberellin-deficient mutants. *Planta,* 171**,** 525-531.

GROOT, S. P. C. & KARSSEN, C. M. 1992. Dormancy and Germination of Abscisic Acid-Deficient Tomato Seeds: Studies with the *sitiens* Mutant. *Plant Physiology,* 99**,** 952-958.

GROSS, B. L. & OLSEN, K. M. 2010. Genetic perspectives on crop domestication. *Trends in plant science,* 15**,** 529-537.

GU, X.-Y., FOLEY, M. E., HORVATH, D. P., ANDERSON, J. V., FENG, J., ZHANG, L., MOWRY, C. R., YE, H., SUTTLE, J. C., KADOWAKI, K.-I. & CHEN, Z. 2011. Association Between Seed Dormancy and Pericarp Color Is Controlled by a Pleiotropic Gene That Regulates Abscisic Acid and Flavonoid Synthesis in Weedy Red Rice. *Genetics,* 189**,** 1515-1524.

GUBLER, F., MILLAR, A. A. & JACOBSEN, J. V. 2005. Dormancy release, ABA and pre-harvest sprouting. *Current opinion in plant biology,* 8**,** 183-187.

GUO, Y., LI, P., YUYAMA, N., TAN, L. B., FU, Y. C., ZHU, Z. F., LIU, F. X., SUN, C. Q. & CAI, H. W. 2015. Identification of Quantitative Trait Locus for Seed Dormancy and Expression Analysis of Four Dormancy-Related Genes in Sorghum. *Tropical Plant Biology,* 8**,** 9-18.

HABERER, G. & MAYER, KLAUS F. X. 2015. Barley: From Brittle to Stable Harvest. *Cell,* 162**,** 469-471.

HACKETT, C. A. 2002. Statistical methods for QTL mapping in cereals. *Plant Molecular Biology,* 48**,** 585-599.

HADAS, A. 1976. WATER-UPTAKE AND GERMINATION OF LEGUMINOUS SEEDS UNDER CHANGING EXTERNAL WATER POTENTIAL IN OSMOTIC SOLUTIONS. *Journal of Experimental Botany,* 27**,** 480-489.

HALL, K. J., PARKER, J. S., ELLIS, T. H. N., TURNER, L., KNOX, M. R., HOFER, J. M. I., LU, J., FERRANDIZ, C., HUNTER, P. J., TAYLOR, J. D. & BAIRD, K. 1997. The relationship between genetic and cytogenetic maps of pea. II. Physical maps of linkage mapping populations. *Genome,* 40**,** 775-769.

HAMMER, K. 1984. Das Domestikationssyndrom. *Die Kulturpflanze,* 32**,** 11-34.

HAN, C. & YANG, P. 2015. Studies on the molecular mechanisms of seed germination. *Proteomics,* 15**,** 1671-1679.

HAN, Y., VIMOLMANGKANG, S., SORIA-GUERRA, R. E., ROSALES-MENDOZA, S., ZHENG, D., LYGIN, A. V. & KORBAN, S. S. 2010. Ectopic Expression of Apple F3'H Genes Contributes to Anthocyanin

Accumulation in the Arabidopsis tt7 Mutant Grown Under Nitrogen Stress. *Plant Physiology,* 153**,** 806-820.
HANE, J. K., MING, Y., KAMPHUIS, L. G., NELSON, M. N., GARG, G., ATKINS, C. A., BAYER, P. E., BRAVO, A., BRINGANS, S., CANNON, S., EDWARDS, D., FOLEY, R., GAO, L.-L., HARRISON, M. J., HUANG, W., HURGOBIN, B., LI, S., LIU, C.-W., MCGRATH, A., MORAHAN, G., MURRAY, J., WELLER, J., JIAN, J. & SINGH, K. B. 2017. A comprehensive draft genome sequence for lupin (Lupinus angustifolius), an emerging health food: insights into plant–microbe interactions and legume evolution. *Plant Biotechnology Journal,* 15**,** 318-330.
HANZAWA, Y., MONEY, T. & BRADLEY, D. 2005. A single amino acid converts a repressor to an activator of flowering. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 7748-7753.
HARLAN, J. R. 1992. *Crops and man,* Madison, Wisconsin, American Society of Agronomy.
HARLAN, J. R., WET, J. & PRICE, E. G. 1973. COMPARATIVE EVOLUTION OF CEREALS. *Evolution,* 27**,** 311-325.
HARMER, S. L. 2009. The Circadian System in Higher Plants. *Annual Review of Plant Biology.* Palo Alto: Annual Reviews.
HARRIS, W. M. 1987. Comparative Ultrastructure of Developing Seed Coats of "Hard-Seeded" and "Soft-Seeded" Varieties of Soybean, Glycine max (L.) Merr. *Botanical Gazette,* 148**,** 324-331.
HASLAM, E. 1993. *Shikimic acid: metabolism and metabolites*, John Wiley & Sons Inc.
HECHT, V., FOUCHER, F., FERRÁNDIZ, C., MACKNIGHT, R., NAVARRO, C., MORIN, J., VARDY, M. E., ELLIS, N., BELTRÁN, J. P., RAMEAU, C. & WELLER, J. L. 2005. Conservation of Arabidopsis Flowering Genes in Model Legumes. *Plant Physiol,* 137**,** 1420-34.
HECHT, V., KNOWLES, C. L., VANDER SCHOOR, J. K., LIEW, L. C., JONES, S. E., LAMBERT, M. J. M. & WELLER, J. L. 2007. Pea LATE BLOOMER1 Is a GIGANTEA Ortholog with Roles in Photoperiodic Flowering, Deetiolation, and Transcriptional Regulation of Circadian Clock Gene Homologs. *Plant Physiol,* 144**,** 648-61.
HECHT, V., LAURIE, R. E., VANDER SCHOOR, J. K., RIDGE, S., KNOWLES, C. L., LIEW, L. C., SUSSMILCH, F. C., MURFET, I. C., MACKNIGHT, R. C. & WELLER, J. L. 2011. The Pea GIGAS Gene Is a FLOWERING LOCUS T Homolog Necessary for Graft-Transmissible Specification of Flowering but Not for Responsiveness to Photoperiod. *The Plant Cell,* 23**,** 147-161.
HELLENS, R. P., MOREAU, C., LIN-WANG, K., SCHWINN, K. E., THOMSON, S. J., FIERS, M. W. E. J., FREW, T. J., MURRAY, S. R., HOFER, J. M. I., JACOBS, J. M. E., DAVIES, K. M., ALLAN, A. C., BENDAHMANE, A., COYNE, C. J., TIMMERMAN-VAUGHAN, G. M. & ELLIS, T. H. N. 2010. Identification of Mendel's White Flower Character. *Plos One,* 5.
HEPWORTH, J. & DEAN, C. 2015. Flowering Locus C's lessons: conserved chromatin switches underpinning developmental timing and adaptation. *Plant physiology,* 168**,** 1237-1245.
HILLMAN, G. C. & DAVIES, M. S. 1990. 6. Domestication rates in wild-type wheats and barley under primitive cultivation. *Biological Journal of the Linnean Society,* 39**,** 39-78.
HIRSCH, A. M. 1992. DEVELOPMENTAL BIOLOGY OF LEGUME NODULATION. *New Phytologist,* 122**,** 211-237.
HISAMATSU, T. & KING, R. W. 2008. The nature of floral signals in Arabidopsis. II. Roles for FLOWERING LOCUS T (FT) and gibberellin. *Journal of experimental botany,* 59**,** 3821-3829.
HOEY, B. K., CROWE, K. R., JONES, V. M. & POLANS, N. O. 1996. A phylogenetic analysis of Pisum based on morphological characters, and allozyme and RAPD markers. *Theoretical and Applied Genetics,* 92**,** 92-100.
HOLDSWORTH, M. J., BENTSINK, L. & SOPPE, W. J. 2008. Molecular networks regulating Arabidopsis seed maturation, after-ripening, dormancy and germination. *New Phytologist,* 179**,** 33-54.
HOLDSWORTH, W. L., GAZAVE, E., CHENG, P., MYERS, J. R., GORE, M. A., COYNE, C. J., MCGEE, R. J. & MAZOUREK, M. 2017. A community resource for exploring and utilizing genetic diversity in the USDA pea single plant plus collection. *Horticulture research* [Online], 4. Available: http://europepmc.org/abstract/MED/28503311

http://europepmc.org/articles/PMC5405346?pdf=render

http://europepmc.org/articles/PMC5405346

https://doi.org/10.1038/hortres.2017.17 [Accessed 2017].

HOLLAND, J. B. 2007. Genetic architecture of complex traits in plants. *Current opinion in plant biology,* 10**,** 156-161.

HRADILOVÁ, I., TRNĚNÝ, O., VÁLKOVÁ, M., CECHOVÁ, M., JANSKÁ, A., PROKEŠOVÁ, L., AAMIR, K., KREZDORN, N., ROTTER, B., WINTER, P., VARSHNEY, R. K., SOUKUP, A., BEDNÁŘ, P., HANÁČEK, P. & SMÝKAL, P. 2017. A Combined Comparative Transcriptomic, Metabolomic, and Anatomical Analyses of Two Key Domestication Traits: Pod Dehiscence and Seed Dormancy in Pea (Pisum sp.). *Frontiers in Plant Science,* 8.

HUFFORD, M. B., XU, X., VAN HEERWAARDEN, J., PYHÄJÄRVI, T., CHIA, J.-M., CARTWRIGHT, R. A., ELSHIRE, R. J., GLAUBITZ, J. C., GUILL, K. E. & KAEPPLER, S. M. 2012. Comparative population genomics of maize domestication and improvement. *Nature genetics,* 44**,** 808.

HUMPHRY, M., LAMBRIDES, C., CHAPMAN, S., AITKEN, E., IMRIE, B., LAWN, R., MCINTYRE, C. & LIU, C. 2005. Relationships between hard-seededness and seed weight in mungbean (Vigna radiata) assessed by QTL analysis. *Plant Breeding,* 124**,** 292-298.

HYTEN, D. L., SONG, Q. J., ZHU, Y. L., CHOI, I. Y., NELSON, R. L., COSTA, J. M., SPECHT, J. E., SHOEMAKER, R. C. & CREGAN, P. B. 2006. Impacts of genetic bottlenecks on soybean genome diversity. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 16666-16671.

ICHINO, T., FUJI, K., UEDA, H., TAKAHASHI, H., KOUMOTO, Y., TAKAGI, J., TAMURA, K., SASAKI, R., AOKI, K., SHIMADA, T. & HARA-NISHIMURA, I. 2014. GFS9/TT9 contributes to intracellular membrane trafficking and flavonoid accumulation in Arabidopsis thaliana. *The Plant Journal,* 80**,** 410-423.

INGRAM, T. J., REID, J. B., MURFET, I. C., GASKIN, P., WILLIS, C. L. & MACMILLAN, J. 1984. INTERNODE LENGTH IN PISUM - THE LE GENE CONTROLS THE 3-BETA-HYDROXYLATION OF GIBBERELLIN-A20 TO GIBBERELLIN-A1. *Planta,* 160**,** 455-463.

ISEMURA, T., KAGA, A., KONISHI, S., ANDO, T., TOMOOKA, N., HAN, O. K. & VAUGHAN, D. A. 2007. Genome dissection of traits related to domestication in azuki bean (Vigna angularis) and comparison with other warm-season legumes. *Annals of Botany,* 100**,** 1053-1071.

ISEMURA, T., KAGA, A., TABATA, S., SOMTA, P., SRINIVES, P., SHIMIZU, T., JO, U., VAUGHAN, D. A. & TOMOOKA, N. 2012. Construction of a Genetic Linkage Map and Genetic Analysis of Domestication Related Traits in Mungbean (Vigna radiata). *PLOS ONE,* 7**,** e41304.

ISEMURA, T., KAGA, A., TOMOOKA, N., SHIMIZU, T. & VAUGHAN, D. A. 2010. The genetics of domestication of rice bean, Vigna umbellata. *Annals of Botany,* 106**,** 927-944.

IŠTVÁNEK, J., DLUHOŠOVÁ, J., DLUHOŠ, P., PÁTKOVÁ, L., NEDĚLNÍK, J. & ŘEPKOVÁ, J. 2017. Gene Classification and Mining of Molecular Markers Useful in Red Clover (Trifolium pratense) Breeding. *Frontiers in Plant Science,* 8.

JAN, I., MICHAL, J., ALEŠ, K. & JANA, Ř. 2014. Genome assembly and annotation for red clover (Trifolium pratense; Fabaceae). *American Journal of Botany,* 101**,** 327-337.

JANG, S. J., SATO, M., SATO, K., JITSUYAMA, Y., FUJINO, K., MORI, H., TAKAHASHI, R., BENITEZ, E. R., LIU, B. H., YAMADA, T. & ABE, J. 2015. A Single-Nucleotide Polymorphism in an Endo-1,4-beta-Glucanase Gene Controls Seed Coat Permeability in Soybean. *Plos One,* 10**,** 19.

JAUDAL, M., ZHANG, L., CHE, C., HURLEY, D. G., THOMSON, G., WEN, J., MYSORE, K. S. & PUTTERILL, J. 2016. MtVRN2 is a Polycomb VRN2-like gene which represses the transition to flowering in the model legume Medicago truncatula. *The Plant Journal,* 86**,** 145-160.

JIGHLY, A., JOUKHADAR, R. & ALAGU, M. 2015. SimpleMap: A Pipeline to Streamline High-Density Linkage Map Construction. *Plant Genome,* 8.

JING, R., AMBROSE, M., KNOX, M., SMYKAL, P., HYBL, M., RAMOS, A., CAMINERO, C., BURSTIN, J., DUC, G. & VAN SOEST, L. 2012. Genetic diversity in European Pisum germplasm collections. *Theoretical and applied genetics,* 125**,** 367-380.

JING, R., VERSHININ, A., GRZEBYTA, J., SHAW, P., SMÝKAL, P., MARSHALL, D., AMBROSE, M. J., ELLIS, T. N. & FLAVELL, A. J. 2010. The genetic diversity and evolution of field pea (Pisum) studied by high throughput retrotransposon based insertion polymorphism (RBIP) marker analysis. *BMC Evolutionary Biology,* 10**,** 44.

JONES, H., LEIGH, F. J., MACKAY, I., BOWER, M. A., SMITH, L. M. J., CHARLES, M. P., JONES, G., JONES, M. K., BROWN, T. A. & POWELL, W. 2008. Population-Based Resequencing Reveals That the Flowering Time Adaptation of Cultivated Barley Originated East of the Fertile Crescent. *Molecular Biology and Evolution,* 25**,** 2211-2219.

KAGA, A., ISEMURA, T., TOMOOKA, N. & VAUGHAN, D. A. 2008. The genetics of domestication of the azuki bean (Vigna angularis). *Genetics,* 178**,** 1013-1036.

KAGA, A., SHIMIZU, T., WATANABE, S., TSUBOKURA, Y., KATAYOSE, Y., HARADA, K., VAUGHAN, D. A. & TOMOOKA, N. 2012. Evaluation of soybean germplasm conserved in NIAS genebank and development of mini core collections. *Breeding Science,* 61**,** 566-592.

KALÓ, P., SERES, A., TAYLOR, S. A., JAKAB, J., KEVEI, Z., KERESZT, A., ENDRE, G., ELLIS, T. H. N. & KISS, G. B. 2004. Comparative mapping between Medicago sativa and Pisum sativum. *Molecular Genetics and Genomics,* 272**,** 235-246.

KANEKO-SUZUKI, M., KURIHARA-ISHIKAWA, R., OKUSHITA-TERAKAWA, C., KOJIMA, C., NAGANO-FUJIWARA, M., OHKI, I., TSUJI, H., SHIMAMOTO, K. & TAOKA, K. I. 2018. TFL1-Like Proteins in Rice Antagonize Rice FT-Like Protein in Inflorescence Development by Competition for Complex Formation with 14-3-3 and FD. *Plant and Cell Physiology,* 59**,** 458-468.

KANG, Y. J., KIM, S. K., KIM, M. Y., LESTARI, P., KIM, K. H., HA, B.-K., JUN, T. H., HWANG, W. J., LEE, T., LEE, J., SHIM, S., YOON, M. Y., JANG, Y. E., HAN, K. S., TAEPRAYOON, P., YOON, N., SOMTA, P., TANYA, P., KIM, K. S., GWAG, J.-G., MOON, J.-K., LEE, Y.-H., PARK, B.-S., BOMBARELY, A., DOYLE, J. J., JACKSON, S. A., SCHAFLEITNER, R., SRINIVES, P., VARSHNEY, R. K. & LEE, S.-H. 2014. Genome sequence of mungbean and insights into evolution within Vigna species. *Nature Communications,* 5**,** 5443.

KANG, Y. J., SATYAWAN, D., SHIM, S., LEE, T., LEE, J., HWANG, W. J., KIM, S. K., LESTARI, P., LAOSATIT, K., KIM, K. H., HA, T. J., CHITIKINENI, A., KIM, M. Y., KO, J.-M., GWAG, J.-G., MOON, J.-K., LEE, Y.-H., PARK, B.-S., VARSHNEY, R. K. & LEE, S.-H. 2015. Draft genome sequence of adzuki bean, Vigna angularis. *Scientific Reports,* 5**,** 8069.

KAZAKOFF, S. H., IMELFORT, M., EDWARDS, D., KOEHORST, J., BISWAS, B., BATLEY, J., SCOTT, P. T. & GRESSHOFF, P. M. 2012. Capturing the biofuel wellhead and powerhouse: the chloroplast and mitochondrial genomes of the leguminous feedstock tree Pongamia pinnata. *PLoS One,* 7**,** e51687.

KAZNOWSKI, L. 1926. Recherches sur le pois.(Pisum). *Mémoires de l'Institut national polonais d'économie rurale á Pulawy,* 7**,** 1-91.

KEARSE, M., MOIR, R., WILSON, A., STONES-HAVAS, S., CHEUNG, M., STURROCK, S., BUXTON, S., COOPER, A., MARKOWITZ, S., DURAN, C., THIERER, T., ASHTON, B., MEINTJES, P. & DRUMMOND, A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics,* 28**,** 1647-1649.

KEVAN, P., GIURFA, M. & CHITTKA, L. 1996. Why are there so many and so few white flowers? *Trends in Plant Science,* 1**,** 252.

KIM, Y. & NIELSEN, R. 2004. Linkage disequilibrium as a signature of selective sweeps. *Genetics,* 167**,** 1513-1524.

KINOSHITA, T., ONO, N., HAYASHI, Y., MORIMOTO, S., NAKAMURA, S., SODA, M., KATO, Y., OHNISHI, M., NAKANO, T., INOUE, S.-I. & SHIMAZAKI, K.-I. 2011. FLOWERING LOCUS T Regulates Stomatal Opening. *Current Biology,* 21**,** 1232-1238.

KNIGHT, A. 1799. XII. An account of some experiments on the fecundation of vegetables. In a letter from Thomas Andrew Knight, Esq. to the Right Hon. Sir Joseph Banks, K. B. P. R. S. *Philosophical Transactions of the Royal Society of London,* 89**,** 195-204.
KNOWPULSE. 2018. *Lentil sequenced genome* [Online]. Available: http://knowpulse.usask.ca/portal/ [Accessed].
KOBAYASHI, Y., KAYA, H., GOTO, K., IWABUCHI, M. & ARAKI, T. 1999. A pair of related genes with antagonistic roles in mediating flowering signals. *Science,* 286**,** 1960-1962.
KOINANGE, E. M. K., SINGH, S. P. & GEPTS, P. 1996. Genetic control of the domestication syndrome in common bean. *Crop Science,* 36**,** 1037-1045.
KONGJAIMUN, A., KAGA, A., TOMOOKA, N., SOMTA, P., VAUGHAN, D. A. & SRINIVES, P. 2012. The genetics of domestication of yardlong bean, Vigna unguiculata (L.) Walp. ssp. unguiculata cv.-gr. sesquipedalis. *Annals of Botany,* 109**,** 1185-1200.
KONZEN, E. R. & TSAI, S. M. 2014. Seed coat shininess in Phaseolus vulgaris: rescuing a neglected trait by its screening on commercial lines and landraces. *Journal of Agricultural Science,* 6**,** 113.
KOSTERIN, O. 2017. Abyssnian pea (Lathyrus schaeferi Kosterin pro Pisum abyssinicum A. Br.)-a problematic taxon. *Acta Biologica Sibirica,* 3.
KOSTERIN, O. E. 1994. Mapping of the third locus for histone H1 genes in pea. *Pisum Genetics,* 24**,** 56-59.
KSIĄŻKIEWICZ, M., RYCHEL, S., NELSON, M. N., WYRWA, K., NAGANOWSKA, B. & WOLKO, B. 2016. Expansion of the phosphatidylethanolamine binding protein family in legumes: a case study of Lupinus angustifolius L. FLOWERING LOCUS T homologs, LanFTc1 and LanFTc2. *BMC Genomics,* 17.
KWON, S.-J., BROWN, A. F., HU, J., MCGEE, R., WATT, C., KISHA, T., TIMMERMAN-VAUGHAN, G., GRUSAK, M., MCPHEE, K. E. & COYNE, C. J. 2012. Genetic diversity, population structure and genome-wide marker-trait association analysis emphasizing seed nutrients of the USDA pea (Pisum sativum L.) core collection. *Genes & Genomics,* 34**,** 305-320.
LADIZINSKY, G. 1985. The genetics of hard seed coat in the genus Lens. *Euphytica,* 34**,** 539-543.
LADIZINSKY, G. 1987. Pulse Domestication before Cultivation. *Economic Botany,* 41**,** 60-65.
LADIZINSKY, G. 1993. Lentil domestication: on the quality of evidence and arguments. *Economic Botany,* 47**,** 60-64.
LANDER, E. S. & BOTSTEIN, D. 1994. MAPPING MENDELIAN FACTORS UNDERLYING QUANTITATIVE TRAITS USING RFLP LINKAGE MAPS (VOL 121, PG 185, 1989). *Genetics,* 136**,** 705-705.
LASHBROOKE, J., COHEN, H., LEVY-SAMOCHA, D., TZFADIA, O., PANIZEL, I., ZEISLER, V., MASSALHA, H., STERN, A., TRAINOTTI, L., SCHREIBER, L., COSTA, F. & AHARONI, A. 2016. MYB107 and MYB9 Homologs Regulate Suberin Deposition in Angiosperms. *Plant Cell,* 28**,** 2097-2116.
LAURIE, R. E., DIWADKAR, P., JAUDAL, M., ZHANG, L., HECHT, V., WEN, J., TADEGE, M., MYSORE, K. S., PUTTERILL, J., WELLER, J. L. & MACKNIGHT, R. C. 2011. The Medicago *FLOWERING LOCUS T* Homolog, *MtFTa1*, Is a Key Regulator of Flowering Time. *Plant Physiology,* 156**,** 2207-2224.
LAVIN, M., HERENDEEN, P. S. & WOJCIECHOWSKI, M. F. 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. *Systematic Biology,* 54**,** 575-594.
LEE, C., YU, D., CHOI, H.-K. & KIM, R. W. 2017. Reconstruction of a composite comparative map composed of ten legume genomes. *Genes & Genomics,* 39**,** 111-119.
LEE, J.-H., YOON, H.-J., TERZAGHI, W., MARTINEZ, C., DAI, M., LI, J., BYUN, M.-O. & DENG, X. W. 2010. DWA1 and DWA2, Two *Arabidopsis* DWD Protein Components of CUL4-Based E3 Ligases, Act Together as Negative Regulators in ABA Signal Transduction. *The Plant Cell,* 22**,** 1716-1732.
LEE, R., BALDWIN, S., KENEL, F., MCCALLUM, J. & MACKNIGHT, R. 2013. FLOWERING LOCUS T genes control onion bulb formation and flowering. *Nature communications,* 4**,** 2884.

LEGESSE, N. & POWELL, A. Relationship between the development of seed coat pigmentation, seed coat adherence to the cotyledons and the rate of imbibition during the maturation of grain legumes. Proceedings of the International Seed Testing Association, 1996.

LEHMENSIEK, A., ECKERMANN, P. J., VERBYLA, A. P., APPELS, R., SUTHERLAND, M. W. & DAGGARD, G. E. 2005. Curation of wheat maps to improve map accuracy and QTL detection. *Australian Journal of Agricultural Research,* 56**,** 1347-1354.

LEI, Y., HANNOUFA, A., PRATES, L. L., SHI, H., WANG, Y., BILIGETU, B., CHRISTENSEN, D. & YU, P. 2018. Effects of TT8 and HB12 Silencing on the Relations between the Molecular Structures of Alfalfa (Medicago sativa) Plants and Their Nutritional Profiles and In Vitro Gas Production. *Journal of Agricultural and Food Chemistry,* 66**,** 5602-5611.

LEJEUNE-HÉNAUT, I., HANOCQ, E., BÉTHENCOURT, L., FONTAINE, V., DELBREIL, B., MORIN, J., PETIT, A., DEVAUX, R., BOILLEAU, M., STEMPNIAK, J.-J., THOMAS, M., LAINÉ, A.-L., FOUCHER, F., BARANGER, A., BURSTIN, J., RAMEAU, C. & GIAUFFRET, C. 2008. The flowering locus Hr colocalizes with a major QTL affecting winter frost tolerance in Pisum sativum L. *Theoretical and Applied Genetics,* 116**,** 1105-1116.

LENSER, T. & THEISSE, G. 2013. Molecular mechanisms involved in convergent crop domestication. *Trends in plant science,* 18**,** 704-714.

LEPINIEC, L., DEBEAUJON, I., ROUTABOUL, J.-M., BAUDRY, A., POURCEL, L., NESI, N. & CABOCHE, M. 2006. Genetics and Biochemistry of Seed Flavonoids. *Annual Review of Plant Biology,* 57**,** 405-430.

LERSTEN, N. R. & GUNN, C. R. 1981. Seed Morphology and Testa Topography in Cicer (Fabaceae: Faboideae). *Systematic Botany,* 6**,** 223-230.

LESTER, D. R., MACKENZIE-HOSE, A. K., DAVIES, P. J., ROSS, J. J. & REID*, J. B. 1999. The influence of the null le-2 mutation on gibberellin levels in developing pea seeds. *Plant Growth Regulation,* 27**,** 83-89.

LESTER, D. R., ROSS, J. J., DAVIES, P. J. & REID, J. B. 1997. Mendel's stem length gene (Le) encodes a gibberellin 3 beta-hydroxylase. *The Plant Cell,* 9**,** 1435-1443.

LEV-YADUN, S., GOPHER, A. & ABBO, S. 2000. The cradle of agriculture. *Science,* 288**,** 1602-1603.

LI, C., ZHANG, B., CHEN, B., JI, L. & YU, H. 2018. Site-specific phosphorylation of TRANSPARENT TESTA GLABRA1 mediates carbon partitioning in Arabidopsis seeds. *Nature communications,* 9**,** 571.

LI, L.-F. & OLSEN, K. 2016. To have and to hold: selection for seed and fruit retention during crop domestication. *Current topics in developmental biology.* Elsevier.

LI, P., CHEN, B., ZHANG, G., CHEN, L., DONG, Q., WEN, J., MYSORE, K. S. & ZHAO, J. 2016. Regulation of anthocyanin and proanthocyanidin biosynthesis by Medicago truncatula bHLH transcription factor MtTT8. *New Phytologist,* 210**,** 905-921.

LI, P., FILIAULT, D., BOX, M. S., KERDAFFREC, E., VAN OOSTERHOUT, C., WILCZEK, A. M., SCHMITT, J., MCMULLAN, M., BERGELSON, J. & NORDBORG, M. 2014a. Multiple FLC haplotypes defined by independent cis-regulatory variation underpin life history diversity in Arabidopsis thaliana. *Genes & Development,* 28**,** 1635-1640.

LI, X., CHEN, L., HONG, M., ZHANG, Y., ZU, F., WEN, J., YI, B., MA, C., SHEN, J. & TU, J. 2012. A large insertion in bHLH transcription factor BrTT8 resulting in yellow seed coat in Brassica rapa. *PLoS One,* 7**,** e44145.

LI, Y., ZHAO, S., MA, J., LI, D., YAN, L., LI, J., QI, X., GUO, X., ZHANG, L., HE, W., CHANG, R., LIANG, Q., GUO, Y., YE, C., WANG, X., TAO, Y., GUAN, R., WANG, J., LIU, Y., JIN, L., ZHANG, X., LIU, Z., ZHANG, L., CHEN, J., WANG, K., NIELSEN, R., LI, R., CHEN, P., LI, W., REIF, J. C., PURUGGANAN, M., WANG, J., ZHANG, M., WANG, J. & QIU, L. 2013. Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. *BMC Genomics,* 14**,** 579.

LI, Y. H., REIF, J. C., JACKSON, S. A., MA, Y. S., CHANG, R. Z. & QIU, L. J. 2014b. Detecting SNPs underlying domestication-related traits in soybean. *Bmc Plant Biology,* 14**,** 8.

LIAN, J., LU, X., YIN, N., MA, L., LU, J., LIU, X., LI, J., LU, J., LEI, B., WANG, R. & CHAI, Y. 2017. Silencing of BnTT1 family genes affects seed flavonoid biosynthesis and alters seed fatty acid composition in Brassica napus. *Plant Science,* 254**,** 32-47.

LIEW, L. C., HECHT, V., SUSSMILCH, F. C. & WELLER, J. L. 2014. The Pea Photoperiod Response Gene *STERILE NODES* Is an Ortholog of *LUX ARRHYTHMO*. *Plant Physiology,* 165**,** 648-657.

LIFSCHITZ, E., AYRE, B. G. & ESHED, Y. 2014. Florigen and anti-florigen – a systemic mechanism for coordinating growth and termination in flowering plants. *Frontiers in Plant Science,* 5**,** 465.

LIM, S., PARK, J., LEE, N., JEONG, J., TOH, S., WATANABE, A., KIM, J., KANG, H., KIM, D. H. & KAWAKAMI, N. 2013. ABA-INSENSITIVE3, ABA-INSENSITIVE5, and DELLAs interact to activate the expression of SOMNUS and other high-temperature-inducible genes in imbibed seeds in Arabidopsis. *The Plant Cell***,** tpc. 113.118604.

LIN, S., VESNA, K., YUANYUAN, Y., LJERKA, K. & GEORGE, H. 2012. Arabidopsis glabra2 mutant seeds deficient in mucilage biosynthesis produce more oil. *The Plant Journal,* 69**,** 37-46.

LIU, B., FUJITA, T., YAN, Z.-H., SAKAMOTO, S., XU, D. & ABE, J. 2007. QTL Mapping of Domestication-related Traits in Soybean (Glycine max). *Annals of Botany,* 100**,** 1027-1038.

LIU, B. H. 2017. *Statistical genomics: linkage, mapping, and QTL analysis*, CRC press.

LIU, X., GUO, L., YOU, J., LIU, X., HE, Y., YUAN, J., LIU, G. & FENG, Z. 2010. Progress of segregation distortion in genetic mapping of plants. *Res J Agron,* 4**,** 78-83.

LIU, X. L., COVINGTON, M. F., FANKHAUSER, C., CHORY, J. & WAGNER, D. R. 2001. ELF3 Encodes a Circadian Clock–Regulated Nuclear Protein That Functions in an Arabidopsis PHYB Signal Transduction Pathway. *The Plant Cell,* 13**,** 1293-1304.

LO, S., MUÑOZ-AMATRIAÍN, M., BOUKAR, O., HERNITER, I., CISSE, N., GUO, Y.-N., ROBERTS, P. A., XU, S., FATOKUN, C. & CLOSE, T. J. 2018. Identification of QTL controlling domestication-related traits in cowpea (Vigna unguiculata L. Walp). *Scientific reports,* 8**,** 6261.

LONG ROWENA L., GORECKI MARTA J., RENTON MICHAEL, SCOTT JOHN K., COLVILLE LOUISE, GOGGIN DANICA E., COMMANDER LUCY E., WESTCOTT DAVID A., CHERRY HILLARY & FINCH-SAVAGE WILLIAM E. 2015. The ecophysiology of seed persistence: a mechanistic view of the journey to germination or demise. *Biological Reviews,* 90**,** 31-59.

LOPEZ-MOLINA, L., MONGRAND, S. & CHUA, N.-H. 2001. A postgermination developmental arrest checkpoint is mediated by abscisic acid and requires the ABI5 transcription factor in Arabidopsis. *Proceedings of the National Academy of Sciences,* 98**,** 4782-4787.

LUSH, W. & EVANS, L. 1980. The seed coats of cowpeas and other grain legumes: structure in relation to function. *Field Crops Research,* 3**,** 267-286.

LUSH, W. M. & EVANS, L. T. 1981. The domestication and improvement of cowpeas (Vigna unguiculata (L.) Walp.). *Euphytica,* 30**,** 579-587.

MA, F., CHOLEWA, E., MOHAMED, T., PETERSON, C. A. & GIJZEN, M. 2004a. Cracks in the palisade cuticle of soybean seed coats correlate with their permeability to water. *Annals of Botany,* 94**,** 213-228.

MA, F. S., CHOLEWA, E., MOHAMED, T., PETERSON, C. A. & GIJZEN, M. 2004b. Cracks in the palisade cuticle of soybean seed coats correlate with their permeability to water. *Annals of Botany,* 94**,** 213-228.

MA, Y., COYNE, C. J., GRUSAK, M. A., MAZOUREK, M., CHENG, P., MAIN, D. & MCGEE, R. J. 2017. Genome-wide SNP identification, linkage map construction and QTL mapping for seed mineral concentrations and contents in pea (Pisum sativum L.). *Bmc Plant Biology,* 17**,** 17.

MAASS, B. L. 2006. Changes in seed Morphology, Dormancy and Germination from wild to Cultivated Hyacinth bean Germplasm (Lablab purpureus: Papilionoideae). *Genetic Resources and Crop Evolution,* 53**,** 1127-1135.

MAASS, B. L. & USONGO, M. F. 2007. Changes in seed characteristics during the domestication of the lablab bean (Lablab purpureus (L.) Sweet : Papilionoideae). *Australian Journal of Agricultural Research,* 58**,** 9-19.

MACAS, J., NEUMANN, P. & NAVRÁTILOVÁ, A. 2007. Repetitive DNA in the pea (Pisum sativum L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and Medicago truncatula. *BMC Genomics,* 8**,** 427.

MACE, E. S., TAI, S., GILDING, E. K., LI, Y., PRENTIS, P. J., BIAN, L., CAMPBELL, B. C., HU, W., INNES, D. J., HAN, X., CRUICKSHANK, A., DAI, C., FRÈRE, C., ZHANG, H., HUNT, C. H., WANG, X., SHATTE, T., WANG, M., SU, Z., LI, J., LIN, X., GODWIN, I. D., JORDAN, D. R. & WANG, J. 2013. Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nature Communications,* 4**,** 2320.

MACKAY, T. F., STONE, E. A. & AYROLES, J. F. 2009. The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics,* 10**,** 565.

MARBACH, I. & MAYER, A. M. 1974. Permeability of Seed Coats to Water as Related to Drying Conditions and Metabolism of Phenolics. *Plant Physiology,* 54**,** 817-820.

MARX, G. 1969. A new seed gene. *Pea New-s Letter,* 1**,** 11-12.

MATILLA, A., GALLARDO, M. & PUGA-HERMIDA, M. I. 2005. Structural, physiological and molecular aspects of heterogeneity in seeds: a review. *Seed Science Research,* 15**,** 63-76.

MATSUBARA, K., OGISO-TANAKA, E., HORI, K., EBANA, K., ANDO, T. & YANO, M. 2012. Natural Variation in Hd17, a Homolog of Arabidopsis ELF3 That is Involved in Rice Photoperiodic Flowering. *Plant and Cell Physiology,* 53**,** 709-716.

MATUS, A., SLINKARD, A. & VANDENBERG, A. 1993. The potential of zero tannin lentil. *New Crops***,** 279-282.

MAUREN, J., C., Y. C., LULU, Z., CHRISTINE, S., S., M. K., PASCAL, R. & JOANNA, P. 2013. Retroelement insertions at the Medicago FTa1 locus in spring mutants eliminate vernalisation but not long-day requirements for early flowering. *The Plant Journal,* 76**,** 580-591.

MAXTED, N. & BENNETT, S. J. 2001. *Plant genetic resources of legumes in the Mediterranean*, Springer Science & Business Media.

MCCALLUM, J., TIMMERMAN-VAUGHAN, G., FREW, T. & RUSSELL, A. 1997. Biochemical and genetic linkage analysis of green seed color in field pea. *Journal of the American Society for Horticultural Science,* 122**,** 218-225.

MCCLUNG, C. R. 2006. Plant Circadian Rhythms. *Plant Cell,* 18**,** 792-803.

MCKEON, G. & MOTT, J. 1982. The effect of temperature on the field softening of hard seed of Stylosanthes humilis and S. hamata in a dry monsoonal climate. *Australian Journal of Agricultural Research,* 33**,** 75-85.

MENDEL, G. 1866. Versuche über Pflanzenhybriden. *Verhandlungen des naturforschenden Vereines in Brunn 4: 3,* 44.

MEYER, R. S., DUVAL, A. E. & JENSEN, H. R. 2012. Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytologist,* 196**,** 29-48.

MEYER, R. S. & PURUGGANAN, M. D. 2013. Evolution of crop species: genetics of domestication and diversification. *Nat Rev Genet,* 14**,** 840-852.

MIAO, Z. H., FORTUNE, J. A. & GALLAGHER, J. 2001. Anatomical structure and nutritive value of lupin seed coats. *Australian Journal of Agricultural Research,* 52**,** 985-993.

MICHAEL, T. P., SALOMÉ, P. A., YU, H. J., SPENCER, T. R., SHARP, E. L., MCPEEK, M. A., ALONSO, J. M., ECKER, J. R. & MCCLUNG, C. R. 2003. Enhanced Fitness Conferred by Naturally Occurring Variation in the Circadian Clock. *Science,* 302**,** 1049-1053.

MICHAELS, S. D. 2009. Flowering time regulation produces much fruit. *Current opinion in plant biology,* 12**,** 75-80.

MICHAELS, S. D. & AMASINO, R. M. 1999. *FLOWERING LOCUS C* Encodes a Novel MADS Domain Protein That Acts as a Repressor of Flowering. *The Plant Cell,* 11**,** 949-956.

MIKIĆ, A., SMÝKAL, P., KENICER, G., VISHNYAKOVA, M., SARUKHANYAN, N., AKOPIAN, J., VANYAN, A., GABRIELYAN, I., SMÝKALOVÁ, I., SHERBAKOVA, E., ZORIĆ, L., ATLAGIĆ, J., ZEREMSKI-ŠKORIĆ, T., ĆUPINA, B., KRSTIĆ, Đ., JAJIĆ, I., ANTANASOVIĆ, S., ĐORĐEVIĆ, V., MIHAILOVIĆ, V.,

IVANOV, A., OCHATT, S. & AMBROSE, M. 2013. The bicentenary of the research on 'beautiful' vavilovia (Vavilovia formosa), a legume crop wild relative with taxonomic and agronomic potential. *Botanical Journal of the Linnean Society,* 172**,** 524-531.
MILLAR, A. A., JACOBSEN, J. V., ROSS, J. J., HELLIWELL, C. A., POOLE, A. T., SCOFIELD, G., REID, J. B. & GUBLER, F. 2006. Seed dormancy and ABA metabolism in Arabidopsis and barley: the role of ABA 8'-hydroxylase. *The Plant Journal,* 45**,** 942-954.
MIR, R. R., KUDAPA, H., SRIKANTH, S., SAXENA, R. K., SHARMA, A., AZAM, S., SAXENA, K., VARMA PENMETSA, R. & VARSHNEY, R. K. 2014. Candidate gene analysis for determinacy in pigeonpea (Cajanus spp.). *Theoretical and Applied Genetics,* 127**,** 2663-2678.
MITCHUM, M. G., YAMAGUCHI, S., HANADA, A., KUWAHARA, A., YOSHIOKA, Y., KATO, T., TABATA, S., KAMIYA, Y. & SUN, T. P. 2006. Distinct and overlapping roles of two gibberellin 3-oxidases in Arabidopsis development. *The Plant Journal,* 45**,** 804-818.
MOÏSE, J. A., HAN, S., GUDYNAITĘ-SAVITCH, L., JOHNSON, D. A. & MIKI, B. L. A. 2005. Seed coats: Structure, development, composition, and biotechnology. *In Vitro Cellular & Developmental Biology - Plant,* 41**,** 620-644.
MOLINA, I., OHLROGGE, J. B. & POLLARD, M. 2008. Deposition and localization of lipid polyester in developing seeds of Brassica napus and Arabidopsis thaliana. *Plant Journal,* 53**,** 437-449.
MURASE, K., HIRANO, Y., SUN, T.-P. & HAKOSHIMA, T. 2008. Gibberellin-induced DELLA recognition by the gibberellin receptor GID1. *Nature,* 456**,** 459.
MURFET, I. 1973. Flowering in Pisum. Hr, a gene for high response to photoperiod. *Heredity,* 31**,** 157.
MURFET, I. 1985. Pisum sativum. *Handbook of flowering,* 4**,** 97-126.
MURFET, I. C. 1971. Flowering in Pisum. Three distinct phenotypic classes determined by the interaction of a dominant early and a dominant late gene. *Heredity,* 26**,** 243.
MURFET, I. C. 1975. Flowering in Pisum: multiple alleles at the lf locus. *Heredity,* 35**,** 85.
MURFET, I. C. & ELLIS, T. H. N. 1998. Evidence for the linkage sequence st…b…Np…le. 28**,** 12-14.
MURFET, I. C. & REID, J. B. 1987. FLOWERING IN PISUM - GIBBERELLINS AND THE FLOWERING GENES. *Journal of Plant Physiology,* 127**,** 23-29.
MURPHY, C. & FULLER, D. Q. 2017. Seed coat thinning during horsegram (Macrotyloma uniflorum) domestication documented through synchrotron tomography of archaeological seeds. *Scientific Reports,* 7**,** 9.
MURRAY, N. J., WILLIAMSON, M. P., LILLEY, T. H. & HASLAM, E. 1994. Study of the interaction between salivary proline-rich proteins and a polyphenol by 1H-NMR spectroscopy. *European Journal of Biochemistry,* 219**,** 923-935.
MUTASA-GÖTTGENS, E. & HEDDEN, P. 2009. Gibberellin as a factor in floral regulatory networks. *Journal of Experimental Botany,* 60**,** 1979-1989.
NAKAMICHI, N. 2014. Adaptation to the local environment by modifications of the photoperiod response in crops. *Plant and Cell Physiology,* 56**,** 594-604.
NAKAMURA, S., POURKHEIRANDISH, M., MORISHIGE, H., SAMERI, M., SATO, K. & KOMATSUDA, T. 2017. Quantitative Trait Loci and Maternal Effects Affecting the Strong Grain Dormancy of Wild Barley (Hordeum vulgare ssp. spontaneum). *Frontiers in Plant Science,* 8.
NAVARRO, C., ABELENDA, J. A., CRUZ-ORÓ, E., CUÉLLAR, C. A., TAMAKI, S., SILVA, J., SHIMAMOTO, K. & PRAT, S. 2011. Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. *Nature,* 478**,** 119.
NAWRATH, C. 2006. Unraveling the complex network of cuticular structure and function. *Current opinion in plant biology,* 9**,** 281-287.
NAWRATH, C., SCHREIBER, L., FRANKE, R. B., GELDNER, N., REINA-PINTO, J. J. & KUNST, L. 2013. Apoplastic Diffusion Barriers in Arabidopsis. *The Arabidopsis Book / American Society of Plant Biologists,* 11**,** e0167.
NELSON, M. N., KSIAZKIEWICZ, M., RYCHEL, S., BESHARAT, N., TAYLOR, C. M., WYRWA, K., JOST, R., ERSKINE, W., COWLING, W. A., BERGER, J. D., BATLEY, J., WELLER, J. L., NAGANOWSKA, B. & WOLKO, B. 2017. The loss of vernalization requirement in narrow-leafed lupin is associated

with a deletion in the promoter and de-repressed expression of a Flowering Locus T (FT) homologue. *New Phytologist,* 213**,** 220-232.
NESBITT, T. C. & TANKSLEY, S. D. 2002. Comparative sequencing in the genus Lycopersicon: Implications for the evolution of fruit size in the domestication of cultivated tomatoes. *Genetics,* 162**,** 365-379.
NESI, N., DEBEAUJON, I., JOND, C., PELLETIER, G., CABOCHE, M. & LEPINIEC, L. 2000. The *TT8* Gene Encodes a Basic Helix-Loop-Helix Domain Protein Required for Expression of *DFR* and *BAN* Genes in Arabidopsis Siliques. *The Plant Cell,* 12**,** 1863-1878.
NEWELL, C. A. & HYMOWITZ, T. 1978. Seed Coat Variation in Glycine Willd. Subgenus Glycine (Leguminosae) by Sem. *Brittonia,* 30**,** 76-88.
NICHOLAS, C. D., LINDSTROM, J. T. & VODKIN, L. O. 1993. Variation of proline rich cell wall proteins in soybean lines with anthocyanin mutations. *Plant Molecular Biology,* 21**,** 145-156.
NITCHER, R., DISTELFELD, A., TAN, C., YAN, L. & DUBCOVSKY, J. 2013. Increased copy number at the HvFT1 locus is associated with accelerated flowering time in barley. *Molecular genetics and genomics : MGG,* 288**,** 261-275.
NYHOLT, D. R. 2000. All LODs Are Not Created Equal. *American Journal of Human Genetics,* 67**,** 282-288.
OCHATT, S., CONREUX, C., SMÝKALOVÁ, I., SMÝKAL, P. & MIKIĆ, A. 2016. Developing biotechnology tools for 'beautiful' vavilovia (Vavilovia formosa), a legume crop wild relative with taxonomic and agronomic potential. *Plant Cell, Tissue and Organ Culture (PCTOC),* 127**,** 637-648.
OH, H. I., HOFF, J. E., ARMSTRONG, G. S. & HAFF, L. A. 1980. Hydrophobic interaction in tannin-protein complexes. *Journal of Agricultural and Food Chemistry,* 28**,** 394-398.
OJOMO, O. A. 1972. Inheritance of seed coat thickness in cowpeas. *Journal of Heredity,* 63**,** 147-149.
OLSEN, K. M. & WENDEL, J. F. 2013a. A bountiful harvest: genomic insights into crop domestication phenotypes. *Annual Review of Plant Biology,* 64**,** 47-70.
OLSEN, K. M. & WENDEL, J. F. 2013b. Crop plants as models for understanding plant adaptation and diversification. *Frontiers in plant science,* 4**,** 290.
ORF, J., CHASE, K., JARVIK, T., MANSUR, L., CREGAN, P., ADLER, F. & LARK, K. 1999. Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. *Crop Science,* 39**,** 1642-1651.
ORTEGA, R. 2018. *The role of the FT genes in the control of flowering in chickpea.* Doctor of Philosophy, University of Tasmania.
OTOBE, K., WATANABE, S. & HARADA, K. 2015. Analysis of QTLs for the micromorphology on the seed coat surface of soybean using recombinant inbred lines. *Seed Science Research,* 25**,** 409-415.
OTOBE, K. & YOSHIOKA, K. 2008. Crop Morphology-Property of the Surface Structure of the Impermeable Seed Coat of Soybean. *Japanese Journal of Crop Science,* 77**,** 69.
PARWEEN, S., NAWAZ, K., ROY, R., POLE, A. K., SURESH, B. V., MISRA, G., JAIN, M., YADAV, G., PARIDA, S. K., TYAGI, A. K., BHATIA, S. & CHATTOPADHYAY, D. 2015. An advanced draft genome assembly of a desi type chickpea (Cicer arietinum L.). *Scientific Reports,* 5**,** 14.
PATERSON, A. H. 2002. What has QTL mapping taught us about plant domestication? *New Phytologist,* 154**,** 591-608.
PATERSON, A. H., LIN, Y.-R., LI, Z., SCHERTZ, K. F., DOEBLEY, J. F., PINSON, S. R. M., LIU, S.-C., STANSEL, J. W. & IRVINE, J. E. 1995. Convergent Domestication of Cereal Crops by Independent Mutations at Corresponding Genetic Loci. *Science,* 269**,** 1714-1718.
PELEMAN, J. D., WYE, C., ZETHOF, J., SØRENSEN, A. P., VERBAKEL, H., VAN OEVEREN, J., GERATS, T. & VAN DER VOORT, J. R. 2005. Quantitative Trait Locus (QTL) Isogenic Recombinant Analysis: A Method for High-Resolution Mapping of QTL Within a Single Population. *Genetics,* 171**,** 1341-1352.
PENG, F. Y., HU, Z. & YANG, R.-C. 2015. Genome-wide comparative analysis of flowering-related genes in Arabidopsis, wheat, and barley. *International journal of plant genomics,* 2015.

PENMETSA, R. V., NOELIA, C.-G., EMILY, M. B., LISA, V., CASTRO, B., KASSA, M. T., SARMA, B. K., DATTA, S., FARMER, A. D., BAEK, J.-M., COYNE, C. J., VARSHNEY, R. K., WETTBERG, E. J. B. & COOK, D. R. 2016. Multiple post-domestication origins of kabuli chickpea through allelic variation in a diversification-associated transcription factor. *New Phytologist,* 211**,** 1440-1451.

PETRUSSA, E., BRAIDOT, E., ZANCANI, M., PERESSON, C., BERTOLINI, A., PATUI, S. & VIANELLO, A. 2013. Plant Flavonoids—Biosynthesis, Transport and Involvement in Stress Responses. *Int J Mol Sci,* 14**,** 14950-73.

PHAN, H. T. T., ELLWOOD, S. R., FORD, R., THOMAS, S. & OLIVER, R. 2006. Differences in syntenic complexity between Medicago truncatula with Lens culinaris and Lupinus albus. *Functional Plant Biology,* 33**,** 775-782.

PICKERSGILL, B. 2018. Parallel vs. Convergent Evolution in Domestication and Diversification of Crops in the Americas. *Frontiers in Ecology and Evolution,* 6.

PIN, P. A. & NILSSON, O. 2012. The multifaceted roles of FLOWERING LOCUS T in plant development. *Plant Cell and Environment,* 35**,** 1742-1755.

PLITMANN, U. & KISLEV, M. 1989. Reproductive changes induced by domestication. *Advances in legume biology. St. Louis: Missouri Botanical Garden,* 487**,** 503.

POETHIG, R. S. 1990. Phase change and the regulation of shoot morphogenesis in plants. *Science,* 250**,** 923-930.

POURCEL, L., ROUTABOUL, J.-M., KERHOAS, L., CABOCHE, M., LEPINIEC, L. & DEBEAUJON, I. 2005. TRANSPARENT TESTA10 Encodes a Laccase-Like Enzyme Involved in Oxidative Polymerization of Flavonoids in Arabidopsis Seed Coat. *The Plant Cell,* 17**,** 2966-2980.

PURUGGANAN, M. D. & FULLER, D. Q. 2009. The nature of selection during plant domestication. *Nature,* 457**,** 843.

PUTTERILL, J., ZHANG, L. L., YEOH, C. C., BALCEROWICZ, M., JAUDAL, M. & GASIC, E. V. 2013. FT genes and regulation of flowering in the legume Medicago truncatula. *Functional Plant Biology,* 40**,** 1199-1207.

QI, S., LIU, K., GAO, C., LI, D., JIN, C., DUAN, S., MA, H., HAI, J. & CHEN, M. 2017. The effect of BnTT8 on accumulation of seed storage reserves and tolerance to abiotic stresses during Arabidopsis seedling establishment. *Plant Growth Regulation,* 82**,** 271-280.

RADCHUK, V. & BORISJUK, L. 2014. Physical, metabolic and developmental functions of the seed coat. *Frontiers in Plant Science,* 5.

RAJANDRAN, V. 2016. *Genetic control of flowering time in lentil.* Doctoral dissertation, University of Tasmania.

RAMEAU, C., BERTHELOOT, J., LEDUC, N., ANDRIEU, B., FOUCHER, F. & SAKR, S. 2014. Multiple pathways regulate shoot branching. *Frontiers in Plant Science,* 5**,** 741.

RAMSAY, L., COMADRAN, J., DRUKA, A., MARSHALL, D. F., THOMAS, W. T., MACAULAY, M., MACKENZIE, K., SIMPSON, C., FULLER, J. & BONAR, N. 2011. INTERMEDIUM-C, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene TEOSINTE BRANCHED 1. *Nature genetics,* 43**,** 169.

RATCLIFFE, O. J., AMAYA, I., VINCENT, C. A., ROTHSTEIN, S., CARPENTER, R., COEN, E. S. & BRADLEY, D. J. 1998. A common mechanism controls the life cycle and architecture of plants. *Development,* 125**,** 1609-1615.

REID, J. B. & ROSS, J. J. 2011. Mendel's Genes: Toward a Full Molecular Characterization. *Genetics,* 189**,** 3-10.

REINECKE, D. M., WICKRAMARATHNA, A. D., OZGA, J. A., KUREPIN, L. V., JIN, A. L., GOOD, A. G. & PHARIS, R. P. 2013. Gibberellin 3-oxidase Gene Expression Patterns Influence Gibberellin Biosynthesis, Growth, and Development in Pea. *Plant Physiology,* 163**,** 929-945.

REMIGEREAU, M.-S., LAKIS, G., REKIMA, S., LEVEUGLE, M., FONTAINE, M. C., LANGIN, T., SARR, A. & ROBERT, T. 2011. Cereal Domestication and Evolution of Branching: Evidence for Soft

Selection in the Tb1 Orthologue of Pearl Millet (Pennisetum glaucum [L.] R. Br.). *PLOS ONE,* 6**,** e22404.
RIDGE, S., DEOKAR, A., LEE, R., DABA, K., MACKNIGHT, R. C., WELLER, J. L. & TAR'AN, B. 2017. The Chickpea Early Flowering 1 (Efl1) Locus Is an Ortholog of Arabidopsis ELF3. *Plant Physiol,* 175**,** 802-15.
ROSS, K. A., ZHANG, L. & ARNTFIELD, S. D. 2010. Understanding water uptake from the induced changes occurred during processing: chemistry of pinto and navy bean seed coats. *International Journal of Food Properties,* 13**,** 631-647.
ROSS-IBARRA, J. 2005. Quantitative trait loci and the study of plant domestication. *Genetica,* 123**,** 197-204.
ROSS-IBARRA, J., MORRELL, P. L. & GAUT, B. S. 2007. Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proceedings of the National Academy of Sciences,* 104**,** 8641-8648.
SADRAS, V. O. 2007. Evolutionary aspects of the trade-off between seed size and number in crops. *Field Crops Research,* 100**,** 125-138.
SAITO, K., YONEKURA-SAKAKIBARA, K., NAKABAYASHI, R., HIGASHI, Y., YAMAZAKI, M., TOHGE, T. & FERNIE, A. R. 2013. The flavonoid biosynthetic pathway in Arabidopsis: Structural and genetic diversity. *Plant Physiology and Biochemistry,* 72**,** 21-34.
SCHAEFER, H., HECHENLEITNER, P., SANTOS-GUERRA, A., DE SEQUEIRA, M. M., PENNINGTON, R. T., KENICER, G. & CARINE, M. A. 2012. Systematics, biogeography, and character evolution of the legume tribe Fabeae with special focus on the middle-Atlantic island lineages. *BMC Evolutionary Biology,* 12**,** 250.
SEDIVY, E. J., WU, F. & HANZAWA, Y. 2017. Soybean domestication: the origin, genetic architecture and molecular bases. *The New phytologist,* 214**,** 539-553.
SEFA-DEDEH, S. & STANLEY, D. 1979. The relationship of microstructure of cowpeas to water absorption and dehulling properties. *Cereal Chem,* 56**,** 379-386.
SHAO, S. Q., MEYER, C. J., MA, F. S., PETERSON, C. A. & BERNARDS, M. A. 2007. The outermost cuticle of soybean seeds: chemical composition and function during imbibition. *Journal of Experimental Botany,* 58**,** 1071-1082.
SHARPE, A. G., RAMSAY, L., SANDERSON, L. A., FEDORUK, M. J., CLARKE, W. E., LI, R., KAGALE, S., VIJAYAN, P., VANDENBERG, A. & BETT, K. E. 2013. Ancient orphan crop joins modern era: gene-based SNP discovery and mapping in lentil. *Bmc Genomics,* 14.
SHEN, Y., XIANG, Y., XU, E., GE, X. & LI, Z. 2018. Major Co-localized QTL for Plant Height, Branch Initiation Height, Stem Diameter, and Flowering Time in an Alien Introgression Derived Brassica napus DH Population. *Frontiers in Plant Science,* 9.
SHI, M.-Z. & XIE, D.-Y. 2014. Biosynthesis and Metabolic Engineering of Anthocyanins in Arabidopsis thaliana. *Recent Patents on Biotechnology,* 8**,** 47-60.
SHIMOMURA, M., KANAMORI, H., KOMATSU, S., NAMIKI, N., MUKAI, Y., KURITA, K., KAMATSUKI, K., IKAWA, H., YANO, R., ISHIMOTO, M., KAGA, A. & KATAYOSE, Y. 2015. The *Glycine max* cv. Enrei Genome for Improvement of Japanese Soybean Cultivars. *International Journal of Genomics,* 2015**,** 358127.
SHIRLEY, B. W. 1996. Flavonoid biosynthesis: 'New' functions for an 'old' pathway. *Trends in Plant Science,* 1**,** 377-382.
SHIRLEY, B. W. 1998. Flavonoids in seeds and grains: physiological function, agronomic importance and the genetics of biosynthesis. *Seed Science Research,* 8**,** 415-422.
SHU, K., LIU, X.-D., XIE, Q. & HE, Z.-H. 2016. Two faces of one seed: hormonal regulation of dormancy and germination. *Molecular plant,* 9**,** 34-45.
SIMON, R. & COUPLAND, G. 1996. Arabidopsisgenes that regulate flowering time in response to day-length. *Seminars in Cell & Developmental Biology,* 7**,** 419-425.
SINDHU, A., RAMSAY, L., SANDERSON, L. A., STONEHOUSE, R., LI, R., CONDIE, J., SHUNMUGAM, A. S. K., LIU, Y., JHA, A. B., DIAPARI, M., BURSTIN, J., AUBERT, G., TAR'AN, B., BETT, K. E.,

WARKENTIN, T. D. & SHARPE, A. G. 2014. Gene-based SNP discovery and genetic mapping in pea. *Theoretical and Applied Genetics,* 127**,** 2225-2241.
SIOL, M., JACQUIN, F., CHABERT-MARTINELLO, M., SMYKAL, P., LE PASLIER, M. C., AUBERT, G. & BURSTIN, J. 2017. Patterns of Genetic Structure and Linkage Disequilibrium in a Large Collection of Pea Germplasm. *G3-Genes Genomes Genetics,* 7**,** 2461-2471.
SLATTERY, H. D., ATWELL, B. J. & KUO, J. 1982. Relationship between Colour, Phenolic Content and Impermeability in the Seed Coat of various Trifolium subterraneum L. genotypes. *Annals of Botany,* 50**,** 373-378.
SMIRNOVA, O. G. 2002. Mapping of gene Inci, accompanied by new translocation, on linkage group III. *Pisum Genetics,* 34**,** 25.
SMITH, B. D. 1997. The initial domestication of Cucurbita pepo in the Americas 10,000 years ago. *Science,* 276**,** 932-934.
SMITH, B. D. 2006. Eastern North America as an independent center of plant domestication. *Proceedings of the National Academy of Sciences,* 103**,** 12223-12228.
SMÝKAL, P., KENICER, G., FLAVELL, A. J., CORANDER, J., KOSTERIN, O., REDDEN, R. J., FORD, R., COYNE, C. J., MAXTED, N. & AMBROSE, M. J. 2011. Phylogeny, phylogeography and genetic diversity of the Pisum genus. *Plant Genetic Resources,* 9**,** 4-18.
SMYKAL, P., VERNOUD, V., BLAIR, M. W., SOUKUP, A. & THOMPSON, R. D. 2014. The role of the testa during development and in establishment of dormancy of the legume seed. *Frontiers in Plant Science,* 5.
SONG, Q., ZHANG, T., STELLY, D. M. & CHEN, Z. J. 2017. Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. *Genome Biology,* 18**,** 99.
SONG, Y. H., SHIM, J. S., KINMONTH-SCHULTZ, H. A. & IMAIZUMI, T. 2015. Photoperiodic Flowering: Time Measurement Mechanisms in Leaves. *In:* MERCHANT, S. S. (ed.) *Annual Review of Plant Biology, Vol 66.* Palo Alto: Annual Reviews.
SOUZA, F. H. & MARCOS-FILHO, J. 2001. The seed coat as a modulator of seed-environment relationships in Fabaceae. *Brazilian Journal of Botany,* 24**,** 365-375.
SOYK, S., MULLER, N. A., PARK, S. J., SCHMALENBACH, I., JIANG, K., HAYAMA, R., ZHANG, L., VAN ECK, J., JIMENEZ-GOMEZ, J. M. & LIPPMAN, Z. B. 2017. Variation in the flowering gene SELF PRUNING 5G promotes day-neutrality and early yield in tomato. *Nature Genetics,* 49**,** 162-168.
SPONSEL, V. M. & REID, J. B. 1992. Use of an Acylcyclohexanedione Growth Retardant, LAB 198 999, to Determine Whether Gibberellin A_{20} Has Biological Activity per se in Dark-Grown Dwarf *le*^{5839} Seedlings of *Pisum sativum*. *Plant Physiology,* 100**,** 651-654.
STAM, P. 1993. CONSTRUCTION OF INTEGRATED GENETIC-LINKAGE MAPS BY MEANS OF A NEW COMPUTER PACKAGE - JOINMAP. *Plant Journal,* 3**,** 739-744.
STETTER, M. G., GATES, D. J., MEI, W. & ROSS-IBARRA, J. 2017. How to make a domesticate. *Current Biology,* 27**,** R896-R900.
SUDHEESH, S., LOMBARDI, M., LEONFORTE, A., COGAN, N. O. I., MATERNE, M., FORSTER, J. W. & KAUR, S. 2015. Consensus Genetic Map Construction for Field Pea (Pisum sativum L.), Trait Dissection of Biotic and Abiotic Stress Tolerance and Development of a Diagnostic Marker for the er1 Powdery Mildew Resistance Gene. *Plant Molecular Biology Reporter,* 33**,** 1391-1403.
SUN, L. J., MIAO, Z. Y., CAI, C. M., ZHANG, D. J., ZHAO, M. X., WU, Y. Y., ZHANG, X. L., SWARM, S. A., ZHOU, L. W., ZHANG, Z. Y. J., NELSON, R. L. & MA, J. X. 2015. GmHs1-1, encoding a calcineurin-like protein, controls hard-seededness in soybean. *Nature Genetics,* 47**,** 939-+.
SUSSMILCH, F. C., BERBEL, A., HECHT, V., VANDER SCHOOR, J. K., FERRÁNDIZ, C., MADUEÑO, F. & WELLER, J. L. 2015. Pea *VEGETATIVE2* Is an *FD* Homolog That Is Essential for Flowering and Compound Inflorescence Development. *The Plant Cell,* 27**,** 1046-1060.

TAKEDA, T., SUWA, Y., SUZUKI, M., KITANO, H., UEGUCHI-TANAKA, M., ASHIKARI, M., MATSUOKA, M. & UEGUCHI, C. 2003. The OsTB1 gene negatively regulates lateral branching in rice. *The Plant Journal,* 33**,** 513-520.

TANAKA, Y., SASAKI, N. & OHMIYA, A. 2008. Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *The Plant Journal,* 54**,** 733-749.

TANG, H., SEZEN, U. & PATERSON, A. H. 2010. Domestication and plant genomes. *Current opinion in plant biology,* 13**,** 160-166.

TANG, H. B., KRISHNAKUMAR, V., BIDWELL, S., ROSEN, B., CHAN, A. N., ZHOU, S. G., GENTZBITTEL, L., CHILDS, K. L., YANDELL, M., GUNDLACH, H., MAYER, K. F. X., SCHWARTZ, D. C. & TOWN, C. D. 2014. An improved genome release (version Mt4.0) for the model legume Medicago truncatula. *Bmc Genomics,* 15.

TANKSLEY, S. D. 1993. Mapping polygenes. *Annual review of genetics,* 27**,** 205-233.

TAYEH, N., ALUOME, C., FALQUE, M., JACQUIN, F., KLEIN, A., CHAUVEAU, A., BÉRARD, A., HOUTIN, H., ROND, C. & KREPLAK, J. 2015. Development of two major resources for pea genomics: the GenoPea 13.2 K SNP Array and a high-density, high-resolution consensus genetic map. *The Plant Journal,* 84**,** 1257-1273.

TAYLOR, S. A. & MURFET, I. C. 1993. Flowering in pea: a mutation from Lfd to lfa and a summary of induced Lf mutations. *Pisum Genetics,* 25**,** 60-63.

TESHOME, A., BRYNGELSSON, T., MENDESIL, E., MARTTILA, S. & GELETA, M. 2016. Enhancing Neoplasm Expression in Field Pea (Pisum sativum) via Intercropping and Its Significance to Pea Weevil (Bruchus pisorum) Management. *Frontiers in Plant Science,* 7.

TIAN, Z. X., WANG, X. B., LEE, R., LI, Y. H., SPECHT, J. E., NELSON, R. L., MCCLEAN, P. E., QIU, L. J. & MA, J. X. 2010. Artificial selection for determinate growth habit in soybean. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 8563-8568.

TOKER, C. 2009. A note on the evolution of kabuli chickpeas as shown by induced mutations in Cicer reticulatum Ladizinsky. *Genetic Resources and Crop Evolution,* 56**,** 7-12.

TSUJI, H. 2017. Molecular function of florigen. *Breeding Science,* 67**,** 327-332.

TURNER, A., BEALES, J., FAURE, S., DUNFORD, R. P. & LAURIE, D. A. 2005. The Pseudo-Response Regulator *Ppd-H1* Provides Adaptation to Photoperiod in Barley. *Science,* 310**,** 1031-1034.

UZHINTSEVA, L. P. & SIDOROVA, K. K. 1988. Genetics of early flowering pea mutants. *Pisum Genetics,* 20**,** 39-40.

VAN OOIJEN, J. 2009. MapQTL® 6, Software for the mapping of quantitative trait in experiment populations of diploid species. *Kyazma BV, Wageningen*.

VAN STADEN, J., MANNING, J. & DICKENS, C. 1987. phylogenetic analysis of the role of plant hormones in the development and germination of legume seeds. *Advances in legume systematics*.

VARSHNEY, R. K., CHEN, W., LI, Y., BHARTI, A. K., SAXENA, R. K., SCHLUETER, J. A., DONOGHUE, M. T. A., AZAM, S., FAN, G., WHALEY, A. M., FARMER, A. D., SHERIDAN, J., IWATA, A., TUTEJA, R., PENMETSA, R. V., WU, W., UPADHYAYA, H. D., YANG, S.-P., SHAH, T., SAXENA, K. B., MICHAEL, T., MCCOMBIE, W. R., YANG, B., ZHANG, G., YANG, H., WANG, J., SPILLANE, C., COOK, D. R., MAY, G. D., XU, X. & JACKSON, S. A. 2011. Draft genome sequence of pigeonpea (Cajanus cajan), an orphan legume crop of resource-poor farmers. *Nature Biotechnology,* 30**,** 83.

VARSHNEY, R. K., MOHAN, S. M., GAUR, P. M., GANGARAO, N., PANDEY, M. K., BOHRA, A., SAWARGAONKAR, S. L., CHITIKINENI, A., KIMURTO, P. K. & JANILA, P. 2013a. Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. *Biotechnology advances,* 31**,** 1120-1134.

VARSHNEY, R. K., ROORKIWAL, M. & NGUYEN, T. 2013b. Legume genomics: from genomic resources to molecular breeding. *The Plant Genome,* 6**,** 1-7.

VARSHNEY, R. K., SAXENA, R. K., UPADHYAYA, H. D., KHAN, A. W., YU, Y., KIM, C., RATHORE, A., KIM, D., KIM, J., AN, S., KUMAR, V., ANURADHA, G., YAMINI, K. N., ZHANG, W., MUNISWAMY, S.,

KIM, J.-S., PENMETSA, R. V., VON WETTBERG, E. & DATTA, S. K. 2017. Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. *Nature Genetics,* 49**,** 1082.

VERSHININ, A. V., ALLNUTT, T. R., KNOX, M. R., AMBROSE, M. J. & ELLIS, T. H. N. 2003. Transposable Elements Reveal the Impact of Introgression, Rather than Transposition, in Pisum Diversity, Evolution, and Domestication. *Molecular Biology and Evolution,* 20**,** 2067-2075.

VLASOVA, A., CAPELLA-GUTIERREZ, S., RENDON-ANAYA, M., HERNANDEZ-ONATE, M., MINOCHE, A. E., ERB, I., CAMARA, F., PRIETO-BARJA, P., CORVELO, A., SANSEVERINO, W., WESTERGAARD, G., DOHM, J. C., PAPPAS, G. J., SABURIDO-ALVAREZ, S., KEDRA, D., GONZALEZ, I., COZZUTO, L., GOMEZ-GARRIDO, J., AGUILAR-MORON, M. A., ANDREU, N., AGUILAR, O. M., GARCIA-MAS, J., ZEHNSDORF, M., VAZQUEZI, M. P., DELGADO-SALINAS, A., DELAYE, L., LOWY, E., MENTABERRY, A., VIANELLO-BRONDANI, R. P., GARCIA, J. L., ALIOTO, T., SANCHEZ, F., HIMMELBAUER, H., SANTALLA, M., NOTREDAME, C., GABALDON, T., HERRERA-ESTRELLA, A. & GUIGO, R. 2016. Genome and transcriptome analysis of the Mesoamerican common bean and the role of gene duplications in establishing tissue and temporal specialization of genes. *Genome Biology,* 17**,** 18.

VOGT, T. 2010. Phenylpropanoid Biosynthesis. *Molecular Plant,* 3**,** 2-20.

VOORRIPS, R. 2002. MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of heredity,* 93**,** 77-78.

VU, D. T., VELUSAMY, V. & PARK, E. 2014. STRUCTURE AND CHEMICAL COMPOSITION OF WILD SOYBEAN SEED COAT RELATED TO ITS PERMEABILITY. *Pakistan Journal of Botany,* 46**,** 1847-1857.

WADE, B. 1937. Breeding and improvement of peas and beans. *Yearbook of agriculture***,** 251.

WAN, L., LI, B., PANDEY, M. K., WU, Y., LEI, Y., YAN, L., DAI, X., JIANG, H., ZHANG, J., WEI, G., VARSHNEY, R. K. & LIAO, B. 2016. Transcriptome Analysis of a New Peanut Seed Coat Mutant for the Physiological Regulatory Mechanism Involved in Seed Coat Cracking and Pigmentation. *Frontiers in Plant Science,* 7.

WANG, J., YANG, J., JIA, Q., ZHU, J., SHANG, Y., HUA, W. & ZHOU, M. 2014. A new QTL for plant height in barley (Hordeum vulgare L.) showing no negative effects on grain yield. *Plos one,* 9**,** e90144.

WANG, R. X., HAI, L., ZHANG, X. Y., YOU, G. X., YAN, C. S. & XIAO, S. H. 2009. QTL mapping for grain filling rate and yield-related traits in RILs of the Chinese winter wheat population Heshangmai × Yu8679. *Theoretical and Applied Genetics,* 118**,** 313-325.

WANG, X., WANG, X., HU, Q., DAI, X., TIAN, H., ZHENG, K., WANG, X., MAO, T., CHEN, J.-G. & WANG, S. 2015. Characterization of an activation-tagged mutant uncovers a role of GLABRA2 in anthocyanin biosynthesis in Arabidopsis. *The Plant Journal,* 83**,** 300-311.

WEBER, H., BORISJUK, L., HEIM, U., SAUER, N. & WOBUS, U. 1997. A role for sugar transporters during seed development: molecular characterization of a hexose and a sucrose carrier in fava bean seeds. *The Plant Cell,* 9**,** 895-908.

WEEDEN, N., MUEHLBAUER, F. & LADIZINSKY, G. 1992. Extensive conservation of linkage relationships between pea and lentil genetic maps. *Journal of Heredity,* 83**,** 123-129.

WEEDEN, N. F. 2006. Fs and U appear to be alleles of a locus near the end of linkage group V. *Pisum Genetics,* 38.

WEEDEN, N. F. 2007. Genetic changes accompanying the domestication of Pisum sativum: Is there a common genetic basis to the 'Domestication syndrome' for legumes? *Annals of Botany,* 100**,** 1017-1025.

WEEDEN, N. F. 2018. Domestication of pea (Pisum sativum L.): the case of the Abyssinicum pea. *Frontiers in plant science,* 9**,** 515.

WEEDEN, N. F., BRAUNER, S. & PRZYBOROWSKI, J. A. 2002. Genetic analysis of pod dehiscence in pea (Pisum sativum L.). *Cellular and Molecular Biology Letters,* 7**,** 657-664.

WEEDEN, N. F., ELLIS, T. H. N., TIMMERMAN-VAUGHAN, G. M., SWIECICKI, W. K., ROZOV, S. M. & BERDNIKOV, V. A. 1998. A consensus linkage map for Pisum sativum. *Pisum Genetic,* 30.

WEEDEN, N. F. & MARX, G. A. 1987. Further genetic analysis and linkage relationships of isozyme loci in the pea: Confirmation of the diploid nature of the genome. *Journal of Heredity,* 78**,** 153-159.

WEEDEN, N. F. & WOLKO, B. 1990. Linkage map for the garden pea (Pisum sativum) based on molecular markers. *Genetic maps,* 5**,** 6.106-6.112.

WEITBRECHT, K., MUELLER, K. & LEUBNER-METZGER, G. 2011. First off the mark: early seed germination. *Journal of Experimental Botany,* 62**,** 3289-3309.

WELLER, J. L. 2007. Update on the genetics of flowering. *Pisum Genetics,* 39.

WELLER, J. L., HECHT, V., LIEW, L. C., SUSSMILCH, F. C., WENDEN, B., KNOWLES, C. L. & VANDER SCHOOR, J. K. 2009. Update on the genetic control of flowering in garden pea. *Journal of Experimental Botany,* 60**,** 2493-2499.

WELLER, J. L., LIEW, L. C., HECHT, V. F. G., RAJANDRAN, V., LAURIE, R. E., RIDGE, S., WENDEN, B., VANDER SCHOOR, J. K., JAMINON, O., BLASSIAU, C., DALMAIS, M., RAMEAU, C., BENDAHMANE, A., MACKNIGHT, R. C. & LEJEUNE-HENAUT, I. 2012. A conserved molecular basis for photoperiod adaptation in two temperate legumes. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** 21158-21163.

WELLER, J. L. & ORTEGA, R. 2015. Genetic control of flowering time in legumes. *Frontiers in Plant Science,* 6.

WELLER, J. L., REID, J. B., TAYLOR, S. A. & MURFET, I. C. 1997. The genetic control of flowering in pea. *Trends in Plant Science,* 2**,** 412-418.

WERKER, E., MARBACH, I. & MAYER, A. M. 1979. Relation Between the Anatomy of the Testa, Water Permeability and the Presence of Phenolics in the Genus Pisum. *Annals of Botany,* 43**,** 765-771.

WESTOBY, M., LEISHMAN, M. & LORD, J. 1996. Comparative ecology of seed size and dispersal. *Phil. Trans. R. Soc. Lond. B,* 351**,** 1309-1318.

WESTON, D. E., ELLIOTT, R. C., LESTER, D. R., RAMEAU, C., REID, J. B., MURFET, I. C. & ROSS, J. J. 2008. The pea DELLA proteins LA and CRY are important regulators of gibberellin synthesis and root growth. *Plant Physiology,* 147**,** 199-205.

WICKLAND, DANIEL P. & HANZAWA, Y. 2015. The FLOWERING LOCUS T/TERMINAL FLOWER 1 Gene Family: Functional Evolution and Molecular Mechanisms. *Molecular Plant,* 8**,** 983-997.

WILLS, D. M. & BURKE, J. M. 2007. Quantitative trait locus analysis of the early domestication of sunflower. *Genetics,* 176**,** 2589-2599.

WILTSHIRE, R. J. E., MURFET, I. C. & REID, J. B. 1994. The genetic control of heterochrony: Evidence from developmental mutants of Pisum sativum L. *Journal of Evolutionary Biology,* 7**,** 447-465.

WOLKO, B. & WEEDEN, N. F. 1990. Additional markers for chromsome 6. *Pisum Genetic,* 22**,** 71-74.

WU, F. Q., SEDIVY, E. J., PRICE, W. B., HAIDER, W. & HANZAWA, Y. 2017. Evolutionary trajectories of duplicated FT homologues and their roles in soybean domestication. *Plant Journal,* 90**,** 941-953.

WU, W. X., ZHENG, X. M., LU, G. W., ZHONG, Z. Z., GAO, H., CHEN, L. P., WU, C. Y., WANG, H. J., WANG, Q., ZHOU, K. N., WANG, J. L., WU, F. Q., ZHANG, X., GUO, X. P., CHENG, Z. J., LEI, C. L., LIN, Q. B., JIANG, L., WANG, H. Y., GE, S. & WAN, J. M. 2013. Association of functional nucleotide polymorphisms at DTH2 with the northward expansion of rice cultivation in Asia. *Proceedings of the National Academy of Sciences of the United States of America,* 110**,** 2775-2780.

WU, Z., IETSWAART, R., LIU, F., YANG, H., HOWARD, M. & DEAN, C. 2016. Quantitative regulation of FLC via coordinated transcriptional initiation and elongation. *Proceedings of the National Academy of Sciences,* 113**,** 218-223.

XIE, D.-Y., SHARMA, S. B., PAIVA, N. L., FERREIRA, D. & DIXON, R. A. 2003. Role of Anthocyanidin Reductase, Encoded by *BANYULS* in Plant Flavonoid Biosynthesis. *Science,* 299**,** 396-399.

XU, W., BOBET, S., LE GOURRIEREC, J., GRAIN, D., DE VOS, D., BERGER, A., SALSAC, F., KELEMEN, Z., BOUCHEREZ, J., ROLLAND, A., MOUILLE, G., ROUTABOUL, J. M., LEPINIEC, L. & DUBOS, C. 2017. TRANSPARENT TESTA 16 and 15 act through different mechanisms to control proanthocyanidin accumulation in Arabidopsis testa. *Journal of Experimental Botany,* 68**,** 2859-2870.

XU, W., DUBOS, C. & LEPINIEC, L. 2015. Transcriptional control of flavonoid biosynthesis by MYB–bHLH–WDR complexes. *Trends in Plant Science,* 20**,** 176-185.

XU, W., GRAIN, D., BOBET, S., LE GOURRIEREC, J., THEVENIN, J., KELEMEN, Z., LEPINIEC, L. & DUBOS, C. 2014. Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB-bHLH-WDR complexes and their targets in Arabidopsis seed. *New Phytologist,* 202**,** 132-144.

XU, W., GRAIN, D., GOURRIEREC, J., HARSCOËT, E., BERGER, A., JAUVION, V., SCAGNELLI, A., BERGER, N., BIDZINSKI, P., KELEMEN, Z., SALSAC, F., BAUDRY, A., ROUTABOUL, J.-M., LEPINIEC, L. & DUBOS, C. 2013. Regulation of flavonoid biosynthesis involves an unexpected complex transcriptional regulation of TT8 expression, in Arabidopsis. *New Phytologist,* 198**,** 59-70.

YAMAGISHI, M., TAKEUCHI, Y., TANAKA, I., KONO, I., MURAI, K. & YANO, M. 2010. Segregation distortion in F 2 and doubled haploid populations of temperate japonica rice. *Journal of genetics,* 89**,** 237-241.

YANG, G., ZHAI, H., WU, H.-Y., ZHANG, X.-Z., LÜ, S.-X., WANG, Y.-Y., LI, Y.-Q., HU, B., WANG, L., WEN, Z.-X., WANG, D.-C., WANG, S.-D., KYUYA, H., XIA, Z.-J. & XIE, F.-T. 2017. QTL effects and epistatic interaction for flowering time and branch number in a soybean mapping population of Japanese×Chinese cultivars. *Journal of Integrative Agriculture,* 16**,** 1900-1912.

YANG, K., JEONG, N., MOON, J.-K., LEE, Y.-H., LEE, S.-H., KIM, H. M., HWANG, C. H., BACK, K., PALMER, R. G. & JEONG, S.-C. 2010. Genetic Analysis of Genes Controlling Natural Variation of Seed Coat and Flower Colors in Soybean. *Journal of Heredity,* 101**,** 757-768.

YEOH, C. C., BALCEROWICZ, M., LAURIE, R., MACKNIGHT, R. & PUTTERILL, J. 2011. Developing a method for customized induction of flowering. *BMC Biotechnology,* 11**,** 36.

YEOH, C. C., BALCEROWICZ, M., ZHANG, L., JAUDAL, M., BROCARD, L., RATET, P. & PUTTERILL, J. 2013. Fine Mapping Links the FTa1 Flowering Time Regulator to the Dominant Spring1 Locus in Medicago. *PLOS ONE,* 8**,** e53467.

YOUNG, N. D. 1994. Constructing a plant genetic linkage map with DNA markers. *DNA-based markers in plants.* Springer.

YU, M., CHEN, G.-Y., PU, Z.-E., ZHANG, L.-Q., LIU, D.-C., LAN, X.-J., WEI, Y.-M. & ZHENG, Y.-L. 2015. Quantitative trait locus mapping for growth duration and its timing components in wheat. *Molecular Breeding,* 35**,** 44.

YU, S., GALVÃO, V. C., ZHANG, Y. C., HORRER, D., ZHANG, T. Q., HAO, Y. H., FENG, Y. Q., WANG, S., SCHMID, M. & WANG, J. W. 2012. Gibberellin regulates the Arabidopsis floral transition through miR156-targeted SQUAMOSA PROMOTER BINDING-LIKE transcription factors. *Plant Cell,* 24**,** 3320-3332.

ZABALA, G. & VODKIN, L. 2003. Cloning of the Pleiotropic *T* Locus in Soybean and Two Recessive Alleles That Differentially Affect Structure and Expression of the Encoded Flavonoid 3′ Hydroxylase. *Genetics,* 163**,** 295-309.

ZAKHRABEKOVA, S., GOUGH, S. P., BRAUMANN, I., MÜLLER, A. H., LUNDQVIST, J., AHMANN, K., DOCKTER, C., MATYSZCZAK, I., KUROWSKA, M., DRUKA, A., WAUGH, R., GRANER, A., STEIN, N., STEUERNAGEL, B., LUNDQVIST, U. & HANSSON, M. 2012. Induced mutations in circadian clock regulator *Mat-a* facilitated short-season adaptation and range extension in cultivated barley. *Proceedings of the National Academy of Sciences,* 109**,** 4326-4331.

ZAYTSEVA, O. O., BOGDANOVA, V. S., MGLINETS, A. V. & KOSTERIN, O. E. 2017. Refinement of the collection of wild peas (Pisum L.) and search for the area of pea domestication with a

deletion in the plastidic psbA-trnH spacer. *Genetic Resources and Crop Evolution,* 64**,** 1417-1430.
ZEDER, M. A. 2015. Core questions in domestication research. *Proceedings of the National Academy of Sciences,* 112**,** 3191-3198.
ZENG, L., COCKS, P., KAILIS, S. & KUO, J. 2005. The role of fractures and lipids in the seed coat in the loss of hardseededness of six Mediterranean legume species. *The Journal of Agricultural Science,* 143**,** 43-55.
ZHANG, X., GARRETON, V. & CHUA, N.-H. 2005. The AIP2 E3 ligase acts as a novel negative regulator of ABA signaling by promoting ABI3 degradation. *Genes & development,* 19**,** 1532-1543.
ZHAO, J. & DIXON, R. A. 2009. MATE Transporters Facilitate Vacuolar Uptake of Epicatechin 3 '-O-Glucoside for Proanthocyanidin Biosynthesis in Medicago truncatula and Arabidopsis. *Plant Cell,* 21**,** 2323-2340.
ZHENG, M., PENG, C., LIU, H., TANG, M., YANG, H., LI, X., LIU, J., SUN, X., WANG, X., XU, J., HUA, W. & WANG, H. 2017. Genome-Wide Association Study Reveals Candidate Genes for Control of Plant Height, Branch Initiation Height and Branch Number in Rapeseed (Brassica napus L.). *Frontiers in Plant Science,* 8.
ZHOU, L., LIU, S., WU, W., CHEN, D., ZHAN, X., ZHU, A., ZHANG, Y., CHENG, S., CAO, L. & LOU, X. 2016. Dissection of genetic architecture of rice plant height and heading date by multiple-strategy-based association studies. *Scientific reports,* 6**,** 29718.
ZHOU, L., ZHANG, J., YAN, J. & SONG, R. 2011. Two transposable element insertions are causative mutations for the major domestication gene teosinte branched 1 in modern maize. *Cell research,* 21**,** 1267.
ZHOU, S., SEKIZAKI, H., YANG, Z., SAWA, S. & PAN, J. 2010. Phenolics in the Seed Coat of Wild Soybean (Glycine soja) and Their Significance for Seed Hardness and Seed Germination. *Journal of Agricultural and Food Chemistry,* 58**,** 10972-10978.
ZHU, G., YE, N. & ZHANG, J. 2009. Glucose-Induced Delay of Seed Germination in Rice is Mediated by the Suppression of ABA Catabolism Rather Than an Enhancement of ABA Biosynthesis. *Plant and Cell Physiology,* 50**,** 644-651.
ZHU, M., LIU, D., LIU, W., LI, D., LIAO, Y., LI, J., FU, C., FU, F., HUANG, H., ZENG, X., MA, X. & WANG, F. 2017. QTL mapping using an ultra-high-density SNP map reveals a major locus for grain yield in an elite rice restorer R998. *Scientific Reports,* 7**,** 10914.
ZOHARY, D. 1989. Pulse domestication and cereal domestication: how different are they? *Economic Botany,* 43**,** 31-34.
ZOHARY, D. & HOPF, M. 1973. Domestication of Pulses in the Old World. *Legumes were companions of wheat and barley when agriculture began in the Near East,* 182**,** 887-894.
ZOHARY, D., HOPF, M. & WEISS, E. 2012. *Domestication of Plants in the Old World: The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin*, Oxford University Press on Demand.

Supplementary Figures

Supplementary Figure 3.1: Table detailing DaRTseq marker sequence and quality.

Access https://cloudstor.aarnet.edu.au/plus/s/E3WKRxNqWi5qXL3

Supplementary Figure 3.2: Table detailing representative markers identified using SimpleMap

Access https://cloudstor.aarnet.edu.au/plus/s/E3WKRxNqWi5qXL3

Supplementary Figure 3.3: Table detailing markers used in linkage map

Access https://cloudstor.aarnet.edu.au/plus/s/E3WKRxNqWi5qXL3

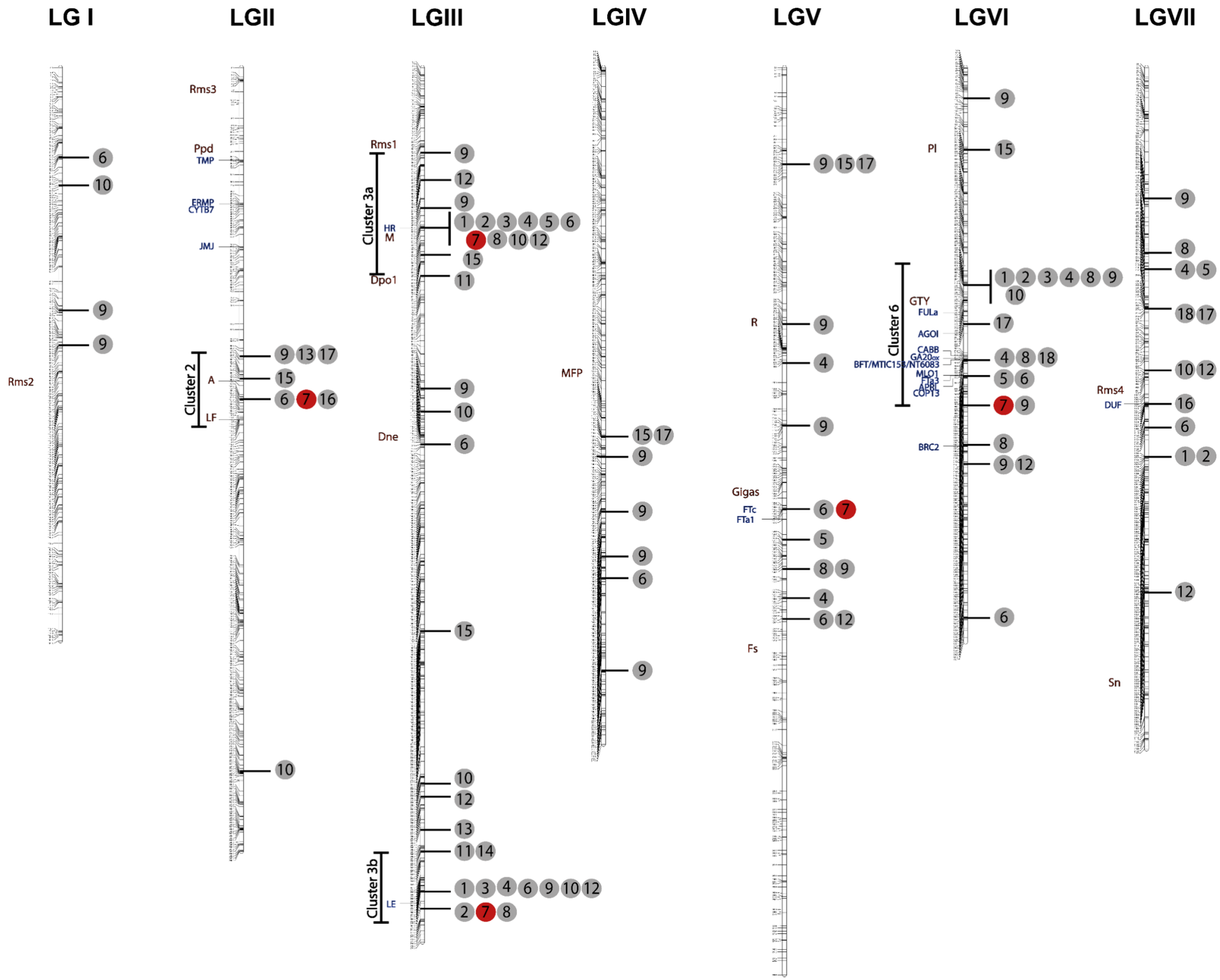
Supplementary Figure 4.1: Phenotypic data collected from RIL population

Access https://cloudstor.aarnet.edu.au/plus/s/E3WKRxNqWi5qXL3

Supplementary Figure 4.2: Overview of genetic control to the 49 measured traits, refer to Table 4.1 for respective trait measurement number. A and B indicate wild and domesticated alleles. PEV A and PEV B is cumulative variation explained by specific allele across population, values in brackets indicates percentage explained from total PEV. Traits highlighted in yellow indicate no major QTL.

Trait	Number of QTLs					Phenotypic variation		
	A	B	Major	Minor	Total	PEV A (%)	PEV B (%)	Total PEV
Total flower colour	1	0	1	0	1	95.7 (100)	0 (0)	95.7
Green testa	2	0	1	1	2	48.5 (100)	0 (0)	48.5
Red testa	2	0	2	0	2	52.9 (100)	0 (0)	52.9
Mottled	2	1	1	2	3	70.4 (96.70)	2.4 (3.3)	72.8
Black spots	2	0	2	0	2	57.5 (100)	0 (0)	57.5
Black hilum	1	0	1	0	1	81.1 (100)	0 (0)	81.1
Permeability	2	0	1	1	2	41.3 (100)	0 (0)	41.3
Testa thickness Digital Calipers	5	0	1	4	5	53.4 (100)	0 (0)	53.4
Testa thickness - Micrometer	4	0	1	3	4	49.7 (100)	0 (0)	49.7
Testa roughness	2	0	1	1	2	71.2 (100)	0 (0)	71.2
Seed desiccation	1	0	1	0	1	21.1 (100)	0 (0)	21.1
Flowering time	3	2	1	4	5	37.2 (74.10)	13 (25.90)	50.2
Floral node	3	1	1	3	4	46.6 (83.21)	9.4 (16.79)	56
Plant height at 22 days	1	1	1	1	2	75 (93.87)	4.9 (6.13)	79.9
Total plant Height	2	2	1	3	4	72 (91.72)	6.5 (8.28)	78.5
Relative branch length at 22 days	0	1	1	0	1	0 (0)	19.9 (100)	19.9
Total relative branch length nodes 1 and 2	2	2	1	3	4	70.6 (91.69)	6.4 (8.31)	77
Internode length between 3 and 6	1	1	1	1	2	63.3 (85.20)	11 (14.80)	74.3
Internode length between 6 and 9	1	2	1	2	3	69.4 (85.05)	12.2 (14.95)	81.6
Internode length between top 3 nodes of plant	1	1	1	1	2	38.1 (73.84)	13.5 (26.16)	51.6
Lower stem thickness	0	4	0	4	4	0 (0)	45.7 (100)	45.7
Upper stem thickness	1	3	2	2	4	20.7 (41.4)	29.3 (58.6)	50
Length of petiole at node 7	1	5	1	5	6	28.9 (53.62)	25 (46.38)	53.9
Length of petiole at node 13	2	2	2	2	4	40.7 (62.04)	24.9 (37.96)	65.6
Primary pod axis	2	2	2	2	4	45.5 (71.88)	17.8 (28.12)	63.3
Secondary pod axis	2	4	1	5	6	69.8 (79.95)	17.5 (20.046)	87.3
Dual pod axis	1	0	1	0	1	34.3 (100)	0 (0)	34.3
Probability of dual pods	1	1	0	2	2	10.4 (48.15)	11.2 (51.85)	21.6
Seed weight	1	1	1	1	2	10.3 (22.69)	35.1 (77.31)	45.4
Seed number	3	2	1	4	5	38.1 (69.15)	17 (30.85)	55.1
Duration of plant growth	3	1	1	3	4	50 (90.58)	5.2 (9.42)	55.2
Number of floral nodes	2	2	2	2	4	36.4 (68.81)	16.5 (31.19)	52.9
Neoplasticity	1	1	1	1	2	69.4 (92.16)	5.9 (7.84)	75.3
Node to change to 4 leaflets	2	1	2	1	3	33.6 (66.40)	17 (33.60)	50.6
Number of nodes expanded at 13 days	1	0	1	0	1	27 (100)	0 (0)	27
Development rate between 7-13 days	2	1	1	2	3	33.4 (82.47)	7.1 (17.53)	40.5
Number of nodes expanded at 41 days	3	1	1	3	4	46.9 (86.21)	7.5 (13.78)	54.4
Development rate between 36 and 41 days	2	2	1	3	4	28.3 (63.60)	16.2 (36.40)	44.5
Number of nodes expanded at 56 days	3	2	1	4	5	57.1 (83.11)	11.6 (16.89)	68.7
Development rate between 49 and 56 days	4	3	2	5	7	50.3 (73.21)	18.4 (26.78)	68.7
Leaf length	1	3	1	3	4	5.7 (11.38)	44.4 (88.62)	50.1
Leaf width	0	2	2	0	2	0 (0)	35.5 (100)	35.5
Relative height of leaf at widest point	0	3	1	2	3	0 (0)	41.9 (100)	41.9
Leaf roundness	0	4	0	4	4	0 (0)	28.9 (100)	28.9
Leaf perimeter	1	2	1	2	3	11.7 (18.66)	46.5 (74.16)	62.7
Leaf area	1	0	0	1	1	12.7 (100)	0 (0)	12.7
Leaf shape principle component 1	1	3	1	3	4	8.5 (22.37)	29.5 (77.63)	38
Leaf shape principle component 2	3	3	2	4	6	60.6 (79.00)	16.1 (20.99)	76.7
Leaf serrations	3	3	2	4	6	32.5 (47.86)	35.4 (52.14)	67.9
Total	85	75	56	104	160			

Supplementary Figure 4.3: Linkage group of pea depicting mapped QTL position of 7 grouped morphological changes. Relative markers positions are mapped on the left of each representative linkage group with anchor marks shown in large print. QTLs are displayed to the right of each linkage group with numbers referring to trait group (Refer to Table 4.1). Red circle indicates flowering time genes. (high resolution version can be accessed on https://cloudstor.aarnet.edu.au/plus/s/E3WKRxNqWi5qXL3)



Supplementary Figure 4.4: Phenotypic data from HR NIL population

Access https://cloudstor.aarnet.edu.au/plus/s/E3WKRxNqWi5qXL3

Supplementary Figure 4.5: Phenotypic data from Clst6 F4 advanced generation segregating population

Access https://cloudstor.aarnet.edu.au/plus/s/E3WKRxNqWi5qXL3

Supplementary Figure 4.6: Details of all seed dormancy and other testa related QTLs. QTLs displayed in order of relative position, where applicable similar measured traits displaying same QTL were grouped together.

Trait code	Trait type	Trait measured	Allelic direction	PEV (%)	LG	Position (cM)	QTL code	QTLs
47	Structure	TT digital caliper	A	16.6	II	91.765	TTd2	*TT2*
48	Structure	TT Micrometer	A	29.1	II	91.765	TTm2	
41	Pigmentation	Green testa	A	6.8	II	98.906	GT2	*A*
42	Pigmentation	red testa	A	27.8	II	99.627	RT2	
43	Pigmentation	Mottled	A	10	II	105.354	Mt2	
44	Pigmentation	Black spots	A	29.4	II	99.627	Bs2	
46	Dormancy	Permeability	A	32.7	II	107.1	Perm2	*Perm2*
42	Pigmentation	red testa	A	25.1	III	60.309	RT3	*RT3*
43	Pigmentation	Mottled	A	60.4	III	60.309	Mt3a	*MT3a*
43	Pigmentation	Mottled	B	2.4	III	179.179	Mt3b	*MT3b*
41	Pigmentation	Green testa	A	41.7	IV	119.542	GT4	*Gt*
47	Structure	TT digital caliper	A	6.7	IV	119.835	TTd4	*TT4*
48	Structure	TT Micrometer	A	9.4	IV	125.904	TTm4	
44	Pigmentation	Black spots/flecking	A	28.1	V	31.54	Bs5	*Bs*
47	Structure	TT digital caliper	A	6.3	V	30.829	TTd5	*TT5*
45	Pigmentation	Black Hilum	A	81.1	VI	26.1	BH6	*Bh*
47	Structure	TT digital caliper	A	12.3	VI	82.22	TTd6	*TT6*
48	Structure	TT Micrometer	A	5.5	VI	81.644	TTm6	
49	Structure	Testa roughness	A	66.6	VI	91.222	GTY	*GTY*
49	Structure	Testa roughness	A	4.6	VII	71.888	Sr7	*Sr7*
47	Structure	TT digital caliper	A	11.5	VII	76.167	TTd7	*TT7*
48	Structure	TT Micrometer	A	5.7	VII	76.707	TTm7	
46	Dormancy	Permeability	A	8.6	VII	107.014	Perm7	*Perm7*

Supplementary Figure 6.1: QTL analysis data for testa thickness, seed coat roughness and permeability within the Clst6 region.

Access https://cloudstor.aarnet.edu.au/plus/s/E3WKRxNqWi5qXL3

Supplementary Figure 6.2: Mapping *GTY* as a qualitative marker in RIL population.

Access https://cloudstor.aarnet.edu.au/plus/s/E3WKRxNqWi5qXL3

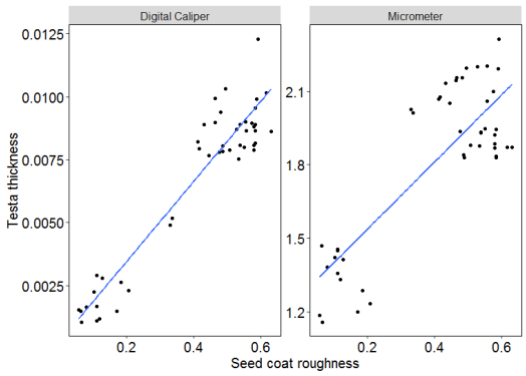
Supplementary Figure 6.3: Table showing M. truncatula orthologous genes in *GTY* locus based on Tayeh et al. (2015) consensus map

No. genes found in *M. truncatula*	Anchor markers	PsCam	Position based from Tayeh et al. (2015) consensus map	orthologous gene ID	Physical map position
1	CABB	PsCam050969	51.1	Medtr6g060175	chr6:20668553-20672457
2		PsCam051457	51.1	Medtr6g065120	chr6:24112211-24113617
3		PsCam052151	51.1	Medtr6g059680	chr6:20475350-20476274
		PsCam053189	51.1		
4		PsCam058525	51.1	Medtr6g060172	chr6:20662407-20668271
		PsCam034349	52.5		
5		PsCam035356	52.5	Medtr6g065580	chr6:24285011-24275638
		PsCam040326	52.5		
6		PsCam049395	52.5	*Medtr2g013540*	*chr2:3653927-3660361*
7		PsCam011150	52.7	Medtr6g066230	chr6:24578667-24584184
		PsCam049193	52.7		
8		PsCam011774	52.8	*Medtr5g025920*	*chr5:10600994-10601488*
9		PsCam012888	52.8	Medtr6g069030	chr6:24827419-24830713
		PsCam023852	52.8		
		PsCam026055	52.8		
		PsCam031070	52.8		
10		PsCam036753	52.8	Medtr6g069010	chr6:24822485-24817427
11		PsCam037027	52.8	Medtr6g066360	chr6:24701161-24699312
12		PsCam045050	52.8	Medtr6g066300	chr6:24666177-24671564
13		PsCam050034	52.8	Medtr6g069040	chr6:24832348-24834380
		PsCam013582	52.9		
14		PsCam034801	52.9	Medtr6g068870	chr6:24756215-24753312
15		PsCam049195	52.9	Medtr6g068920	chr6:24768282-24773470
16		PsCam052421	52.9	Medtr6g068850	chr6:24744886-24747929
		PsCam014621	53		
		PsCam014693	53		
17		PsCam017079	53	Medtr6g069050	chr6:24840126-24835251
18		PsCam044238	53	Medtr6g078450	chr6:29462673-29455806
		PsCam003215	53.1		
		PsCam005025	53.2		
19		PsCam002510	53.3	Medtr6g078490	chr6:29481594-29474654
20		PsCam035162	53.3	*Medtr5g031990*	*chr5:13717562-13717792*
21		PsCam037327	53.3	Medtr6g078330	chr6:29404288-29408831
22		PsCam009675	54	Medtr6g077750	chr6:29197410-29191150
		PsCam021080	54		
23		PsCam034141	54	Medtr6g077740	chr6:29186028-29189722
24		PsCam042544	54	Medtr6g077830	chr6:29228939-29221801
25		PsCam049828	54	Medtr6g077860	chr6:29234964-29244137

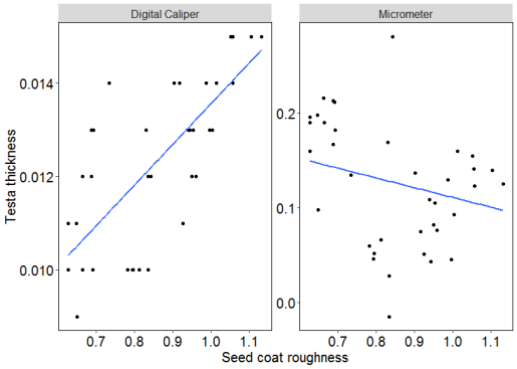
26		PsCam050194	54	Medtr6g079660	chr6:29164021-29170847
27		PsCam051070	54	Medtr6g077820	chr6:29221006-29215746
28		PsCam002015	54.3	Medtr6g079650	chr6:29053437-29053120
29		PsCam017057	54.3	Medtr6g081020	chr6:28892983-28885084
30		PsCam039697	55	Medtr6g074860	chr6:27795423-27791872
		PsCam049478	55.1		
31		PsCam000343	55.4	Medtr6g477860	chr6:28638569-28642957
32		PsCam001042	55.9	*Medtr4g072350*	*chr4:27439852-27441525*
33		PsCam004588	55.9	Medtr6g478000	chr6:28772794-28779662
		PsCam009542	55.9		
34		PsCam023306	55.9	Medtr6g074905	chr6:27734145-27726766
35		PsCam030397	55.9	Medtr6g075290	chr6:27891022-27898162
36		PsCam033992	55.9	*Medtr5g007630*	*chr5:1434626-1440207*
37		PsCam036099	55.9	Medtr6g477820	chr6:28597327-28601831
38	AGO1	PsCam043936	55.9	Medtr6g477980	chr6:28760101-28768915
39		PsCam044932	55.9	*Medtr4g016590*	*chr4:5074660-5083977*
40		PsCam049563	55.9	Medtr6g477780	chr6:28543672-28545494
41		PsCam000836	56.3	Medtr6g072020	chr6:26695311-26691594
42		PsCam000926	56.3	Medtr6g071625	chr6:26581301-26568910
		PsCam001177	56.3		
43		PsCam023444	56.3	Medtr6g072010	chr6:26687163-26691200
		PsCam028836	56.3		
		PsCam037268	56.3		
44		PsCam045473	56.6	Medtr2g059350	chr2:24497012-24493288
45		PsCam038798	56.8	Medtr2g461480	chr2:25388385-25385593
46		PsCam060066	56.8	Medtr2g059590	chr2:24591839-24599933
		PsCam004102	56.9		
47		PsCam025564	57.2	Medtr2g461920	chr2:25606881-25611792
48	FULa	PsCam034575	57.2		

Supplementary Figure 6.4: Correlation between testa thickness and seed coat roughness around 40cM region around peak marker for (a) *TT2* (b) *TT4* (c) *TT5* in RIL population. P value <0.01 except for *TT5* using micrometer (P =0.72).

a)



b)



c)